GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:05:42 ; Search time 11154 Seconds
(without alignments)
11391.904 Million cell updates/sec

Title: US-09-900-751-1
Perfect score: 3106
Sequence: 1 catggtagacggctgcccgg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3106	100.0	3106	6	AX395268	AX395268 Sequence
2	3106	100.0	3106	10	AF042822	AF042822 Mus muscu
3	3102.8	99.9	3248	10	BC005496	BC005496 Mus muscu
4	2595	83.5	3174	10	AB049189	AB049189 Rattus no
5	2313.6	74.5	2568	10	AB037898	AB037898 Rattus no
6	1905.6	61.4	3273	9	BC030532	BC030532 Homo sapi
7	1889.8	60.8	3149	9	AF118224	AF118224 Homo sapi
8	1884.8	60.7	3120	9	AF133086	AF133086 Homo sapi
9	1884.6	60.7	3128	9	AB030036	AB030036 Homo sapi
10	1883.2	60.6	3147	6	AR081724	AR081724 Sequence
11	1883.2	60.6	3147	6	AR229704	AR229704 Sequence
c 12	1883.2	60.6	3147	6	AR229712	AR229712 Sequence
13	1883.2	60.6	3147	6	AX207897	AX207897 Sequence
14	1883.2	60.6	3147	6	AX207945	AX207945 Sequence
15	1883.2	60.6	3147	9	AF057145	AF057145 Homo sapi
16	1530.8	49.3	2900	6	AR081725	AR081725 Sequence
17	1530.8	49.3	2900	6	AR229705	AR229705 Sequence
18	1530.8	49.3	2900	9	HSU20428	U20428 Human SNC19
19	1175.8	37.9	2148	9	BC005826	BC005826 Homo sapi
20	945.4	30.4	1823	9	BC018146	BC018146 Homo sapi
21	809.6	26.1	3487	5	AB038498	AB038498 Xenopus l
22	701.2	22.6	1553	6	AR263832	AR263832 Sequence
c 23	623.2	20.1	225649	2	AC114542	AC114542 Mus muscu
24	556.4	17.9	726	6	AX473052	AX473052 Sequence
25	551.8	17.8	723	6	E13204	E13204 Human cDNA
c 26	359.6	11.6	236387	2	AC096042	AC096042 Rattus no
27	359.6	11.6	261906	2	AC112632	AC112632 Rattus no
28	359.6	11.6	277797	2	AC128347	AC128347 Rattus no
29	306.8	9.9	434	6	AR273737	AR273737 Sequence
30	306.8	9.9	434	6	AR277318	AR277318 Sequence
31	306.8	9.9	434	6	AX368770	AX368770 Sequence
32	265	8.5	591	6	AX261852	AX261852 Sequence
33	244.8	7.9	3183	10	BC029645	BC029645 Mus muscu
34	242.4	7.8	3194	9	AY055384	AY055384 Homo sapi
35	240.8	7.8	2449	9	HSA319876	AJ319876 Homo sapi
36	240.8	7.8	2672	6	AX207905	AX207905 Sequence
37	240.8	7.8	3104	6	AX207903	AX207903 Sequence
38	239.2	7.7	1327	6	AX538208	AX538208 Sequence
39	239.2	7.7	2409	6	AX360098	AX360098 Sequence
40	239.2	7.7	3143	6	AX538206	AX538206 Sequence
c 41	238	7.7	240079	2	AC128364	AC128364 Rattus no
42	212.2	6.8	174526	9	AP001183	AP001183 Homo sapi
c 43	212.2	6.8	182736	2	AC021672	AC021672 Homo sapi
44	212.2	6.8	190314	9	AC019227	AC019227 Homo sapi
45	206	6.6	734	6	AX375702	AX375702 Sequence

ALIGNMENTS

RESULT 1
AX395268
LOCUS AX395268 3106 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 5 from Patent WO0203787.
ACCESSION AX395268
VERSION AX395268.1 GI:21066293
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Allen,K.D. and Leviten,M.W.
TITLE Transgenic mice containing targeted gene disruptions
JOURNAL Patent: WO 0203787-A 5 17-JAN-2002;

Deltagen, Inc. (US)

FEATURES Location/Qualifiers
source 1. .3106
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

BASE COUNT 692 a 862 c 897 g 655 t

ORIGIN

Query Match 100.0%; Score 3106; DB 6; Length 3106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240

Qy        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300

Qy        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360

Qy        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420

Qy        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480

Qy        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540

Qy        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600

Qy        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660

Qy        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720

Qy        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780

Qy        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840

Qy        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900

Qy        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960

Qy        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020

Qy       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080

Qy       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140

Qy       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200

Qy       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260

Qy       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320

Qy       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380

Qy       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440

Qy       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500

Qy       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560

Qy       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620

Qy       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680

Qy       1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740

Qy       1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800

Qy       1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860

Qy       1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920

Qy       1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980

Qy       1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
```

```
Db      1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040

Qy      2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100

Db      2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100

Qy      2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160

Db      2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160

Qy      2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220

Db      2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220

Qy      2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280

Db      2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280

Qy      2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340

Db      2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340

Qy      2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400

Db      2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400

Qy      2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460

Db      2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460

Qy      2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520

Db      2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520

Qy      2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580

Db      2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580

Qy      2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640

Db      2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640

Qy      2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700

Db      2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700

Qy      2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2760

Db      2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCAGACTGTGAACT 2760

Qy      2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820

Db      2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820

Qy      2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880

Db      2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880

Qy      2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940

Db      2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940

Qy      2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000

Db      2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000

Qy      3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060

Db      3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060

Qy      3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106

Db      3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
```

```
RESULT 2
AF042822
LOCUS       AF042822               3106 bp    mRNA    linear   ROD 27-MAR-2000
DEFINITION  Mus musculus epithin mRNA, complete cds.
ACCESSION   AF042822
VERSION     AF042822.3  GI:7330637
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3106)
  AUTHORS   Kim,M.G., Chen,C., Lyu,M.S., Cho,E.G., Park,D., Kozak,C. and
            Schwartz,R.H.
  TITLE     Cloning and chromosomal mapping of a gene isolated from thymic
            stromal cells encoding a new mouse type II membrane serine
            protease, epithin, containing four LDL receptor modules and two CUB
            domains
  JOURNAL   Immunogenetics 49 (5), 420-428 (1999)
  MEDLINE   99216440
   PUBMED   10199918
REFERENCE   2  (bases 1 to 3106)
  AUTHORS   Kim,M.G., Chen,C., Cho,E., Park,D. and Schwartz,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Laboratory of Cellular and Molecular
            Immunology, NIAID, NIH, 9000 Rockville Pike, Bethesda, MD 20892,
            USA
REFERENCE   3  (bases 1 to 3106)
  AUTHORS   Kim,M.G., Chen,C., Cho,E., Park,D. and Schwartz,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Laboratory of Cellular and Molecular
            Immunology, NIAID, NIH, 9000 Rockville Pike, Bethesda, MD 20892,
            USA
  REMARK    Sequence update by submitter
REFERENCE   4  (bases 1 to 3106)
  AUTHORS   Kim,M.G., Chen,C., Cho,E., Park,D. and Schwartz,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2000) Laboratory of Cellular and Molecular
            Immunology, NIAID, NIH, 9000 Rockville Pike, Bethesda, MD 20892,
            USA
  REMARK    Sequence update by submitter
COMMENT     On Mar 27, 2000 this sequence version replaced gi:6435480.
FEATURES             Location/Qualifiers
     source          1. .3106
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C.B.17scid"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /cell_type="stromal"
                     /tissue_type="thymus"
                     /note="severe combined immunodeficient (SCID)"
     CDS             63. .2630
                     /function="serine protease"
                     /codon_start=1
                     /product="epithin"
                     /protein_id="AAD02230.3"
                     /db_xref="GI:7330638"
                     /translation="MGSNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPANNA
                     KKVEKRGPRRWVVLVAVLFSFLLLSLMAGLLVWHFHYRNVRVQKVFNGHLRITNEIFL
                     DAYENSTSTEFISLASQVKEALKLLYNEVPVLGPYHKKSAVTAFSEGSVIAYYWSEFS
                     IPPHLAEEVDRAMAVERVVTLPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFAL
                     HAHGAAVTRFTTPGFPNSPYPAHARCQWVLRGDADSVLSLTFRSFDVAPCDEHGSDLV
                     TVYDSLSPMEPHAVVRLCGTFSPSYNLTFLSSQNVFLVTLITNTDRRHPGFEATFFQL
                     PKMSSCGGFLSDTQGTFSSPYYPGHYPPNINCTWNIKVPNNRNVKVRFKLFYLVDPNV
                     PVGSCTKDYVEINGEKYCGERSQFVVSSNSSKITVHFHSDHSYTDTGFLAEYLSYDSN
                     DPCPGMFMCKTGRCIRKELRCDGWADCPDYSDERYCRCNATHQFTCKNQFCKPLFWVC
                     DSVNDCGDGSDEEGCSCPAGSFKCSNGKCLPQSQKCNGKDNCGDGSDEASCDSVNVVS
                     CTKYTYRCQNGLCLSKGNPECDGKTDCSDGSDEKNCDCGLRSFTKQARVVGGTNADEG
                     EWPWQVSLHALGQGHLCGASLISPDWLVSAAHCFQDDKNFKYSDYTMWTAFLGLLDQS
```

KRSASGVQELKLKRIITHPSFNDFTFDYDIALLELEKSVEYSTVVRPICLPDATHVFP
AGKAIWVTGWGHTKEGGTGALILQKGEIRVINQTTCEDLMPQQITPRMMCVGFLSGGV
DSCQGDSGGPLSSAEKDGRMFQAGVVSWGEGCAQRNKPGVYTRLPVVRDWIKEHTGV"

BASE COUNT 692 a 862 c 897 g 655 t
ORIGIN

Query Match 100.0%; Score 3106; DB 10; Length 3106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CATGGTAGACGGCTGCCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240

Qy        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300

Qy        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360

Qy        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420

Qy        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480

Qy        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540

Qy        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600

Qy        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660

Qy        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720

Qy        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780

Qy        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840

Qy        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900

Qy        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960

Qy        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020

Qy       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080

Qy       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140

Qy       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200

Qy       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260

Qy       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320

Qy       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380

Qy       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440

Qy       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500

Qy       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560

Qy       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620

Qy       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680

Qy       1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740

Qy       1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800

Qy       1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860

Qy       1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920

Qy       1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980

Qy       1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
```

```
Qy       2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100
Db       2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100

Qy       2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160
Db       2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160

Qy       2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220
Db       2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220

Qy       2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280
Db       2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280

Qy       2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
Db       2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340

Qy       2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400
Db       2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400

Qy       2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
Db       2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460

Qy       2461 CCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
Db       2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520

Qy       2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
Db       2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580

Qy       2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
Db       2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640

Qy       2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
Db       2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700

Qy       2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
Db       2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCAGACTGTGAACT 2760

Qy       2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820
Db       2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820

Qy       2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880
Db       2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880

Qy       2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940
Db       2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940

Qy       2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000
Db       2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000

Qy       3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060
Db       3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060

Qy       3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
Db       3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
```

```
RESULT 3
BC005496
LOCUS       BC005496               3248 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus suppression of tumorigenicity 14 (colon carcinoma),
            mRNA (cDNA clone MGC:7395 IMAGE:3488059), complete cds.
ACCESSION   BC005496
VERSION     BC005496.1  GI:13529565
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3248)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 3248)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 7 Row: i Column: 16
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .3248
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="MGC:7395 IMAGE:3488059"
                     /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
                     old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam5"
                     /lab_host="DH10B"
```

```
                     /note="Vector: pCMV-SPORT6"
     gene            1. .3248
                     /gene="St14"
                     /note="synonym: Epithin"
                     /db_xref="LocusID:19143"
                     /db_xref="MGI:1338881"
     CDS             173. .2740
                     /codon_start=1
                     /product="suppression of tumorigenicity 14 (colon
                     carcinoma)"
                     /protein_id="AAH05496.1"
                     /db_xref="GI:13529566"
                     /db_xref="LocusID:19143"
                     /translation="MGSNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPANNA
                     KKVEKRGPRRWVVLVAVLFSFLLLSLMAGLLVWHFHYRNVRVQKVFNGHLRITNEIFL
                     DAYENSTSTEFISLASQVKEALKLLYNEVPVLGPYHKKSAVTAFSEGSVIAYYWSEFS
                     IPPHLAEEVDRAMAVERVVTLPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFAL
                     HAHGAAVTRFTTPGFPNSPYPAHARCQWVLRGDADSVLSLTFRSFDVAPCDEHGSDLV
                     TVYDSLSPMEPHAVVRLCGTFSPSYNLTFLSSQNVFLVTLITNTDRRHPGFEATFFQL
                     PKMSSCGGFLSDTQGTFSSPYYPGHYPPNINCTWNIKVPNNRNVKVRFKLFYLVDPNV
                     PVGSCTKDYVEINGEKYCGERSQFVVSSNSSKITVHFHSDHSYTDTGFLAEYLSYDSN
                     DPCPGMFMCKTGRCIRKELRCDGWADCPDYSDERYCRCNATHQFTCKNQFCKPLFWVC
                     DSVNDCGDGSDEEGCSCPAGSFKCSNGKCLPQSQKCNGKDNCGDGSDEASCDSVNVVS
                     CTKYTYRCQNGLCLSKGNPECDGKTDCSDGSDEKNCDCGLRSFTKQARVVGGTNADEG
                     EWPWQVSLHALGQGHLCGASLISPDWLVSAAHCFQDDKNFKYSDYTMWTAFLGLLDQS
                     KRSASGVQELKLKRIITHPSFNDFTFDYDIALLELEKSVEYSTVVRPICLPDATHVFP
                     AGKAIWVTGWGHTKEGGTGALILQKGEIRVINQTTCEDLMPQQITPRMMCVGFLSGGV
                     DSCQGDSGGPLSSAEKDGRMFQAGVVSWGEGCAQRNKPGVYTRLPVVRDWIKEHTGV"
BASE COUNT      734 a    918 c    928 g    668 t
ORIGIN
```

Query Match 99.9%; Score 3102.8; DB 10; Length 3248;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3104; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        111 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 170

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 230

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        231 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 290

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        291 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 350

Qy        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        351 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 410

Qy        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        411 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 470

Qy        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        471 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 530

Qy        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        531 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 590

Qy        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 650

Qy        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 710

Qy        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 770

Qy        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 830

Qy        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 890

Qy        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 950

Qy        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 1010

Qy        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 1070

Qy        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1130

Qy       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1190

Qy       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1250

Qy       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1251 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1310

Qy       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1311 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1370

Qy       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1371 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1430

Qy       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1431 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1490

Qy       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1491 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1550

Qy       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1551 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1610

Qy       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1611 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1670

Qy       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1671 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1730

Qy       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1731 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1790
```

```
Qy      1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1791 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1850

Qy      1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1851 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1910

Qy      1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1911 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1970

Qy      1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1971 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 2030

Qy      1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2031 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 2090

Qy      1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2091 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2150

Qy      2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2151 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2210

Qy      2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2211 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2270

Qy      2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2271 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2330

Qy      2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2331 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2390

Qy      2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2391 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2450

Qy      2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2451 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2510

Qy      2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2511 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2570

Qy      2461 CCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2571 CCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2630

Qy      2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2631 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2690

Qy      2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2691 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2750

Qy      2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2751 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2810

Qy      2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2811 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2870

Qy      2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2871 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2930

Qy      2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2931 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2990

Qy      2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2991 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 3050

Qy      2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3051 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3110

Qy      3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3111 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3170

Qy      3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
             |||||||||||||||||||||||  ||||||||||||||||||
Db      3171 TTGAGAGTAAAACATTTTATTTCTTTAAAAAAAAAAAAAAAAAAA 3216
```

```
RESULT 4
AB049189
LOCUS       AB049189               3174 bp    mRNA    linear   ROD 14-APR-2001
DEFINITION  Rattus norvegicus mRNA for membrane bound arginine specific serine
            protease, complete cds.
ACCESSION   AB049189
VERSION     AB049189.1  GI:10336526
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Inoue,H., Takahashi,K. and Kishi,K.
  TITLE     membrane-bound arginine specific serine protease
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 3174)
  AUTHORS   Inoue,H., Takahashi,K. and Kishi,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2000) Hideshi Inoue, Tokyo University of Pharmacy
            and Life Science, School of Life Science; 1432-1 Horinouchi,
            Hachioji-shi, Tokyo 192-0392, Japan (E-mail:hinoue@ls.toyaku.ac.jp,
            Tel:81-426-76-7153, Fax:81-426-76-7157)
FEATURES             Location/Qualifiers
     source          1. .3174
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /sex="male"
                     /tissue_type="duodenum"
                     /dev_stage="7 weeks old"
     CDS             127. .2694
                     /codon_start=1
                     /evidence=experimental
                     /product="membrane bound arginine specific serine
                     protease"
                     /protein_id="BAB13765.1"
                     /db_xref="GI:10336527"
                     /translation="MGNNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPVNNA
                     KQVEKRGPRRWVVMVAVVFSFLLLSLMAGLLVWHFHYRNVRIQKVFNGHLRITNENFL
                     DAYENSTSTEFISLASQVKEALKLMYSEVPVLGPYHKKSTVTAFSEGSVIAYYWSEFS
                     IPPHLEEEVDRAMAVERVVTLPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFAL
                     HARGRTVTRFTTPGFPNSPYPAHARCQWVLRGDADSVLSLTFRSFDVAPCDGHDSDLV
                     TVYDSLSPMEPHAVVRLCGTFSPSYNLTFLSSQNVFLVTLITNTDRRHPGFEATFFQL
                     PKMSSCGGLLSEAQGTFSSPYYPGHYPPNINCTWNIKVPNNRNVKVRFKLFYLVDPNI
                     PVGSCTKDYVEINGEKFCGERSQFVVSSNSSKITVHFHSDHSYTDTGFLAEYLSYDSN
```

```
                     DPCPGMFMCKTGRCIRKDLRCDGWADCPDYSDERHCRCNATHQFMCKNQFCKPLFWVC
                     DSVNDCGDGSDEEGCSCPAGSFKCSNGKCLPQSQQCNGKDDCGDGSDEASCDNVNAVS
                     CTKYTYRCQNGLCLNKGNPECDGKKDCSDGSDEKNCDCGLRSFTKQARVVGGTNADEG
                     EWPWQVSLHALGQGHLCGASLISPDWLVSAAHCFQDETIFKYSDHTMWTAFLGLLDQS
                     KRSASGVQEHKLKRIITHPSFNDFTFDYDIALLELEKPAEYSTVVRPICLPDNTHVFP
                     AGKAIWVTGWGHTKEGGTGALILQKGEIRVINQTTCEELLPQQITPRMMCVGFLSGGV
                     DSCQGDSGGPLSSVEKDGRIFQAGVVSWGEGCAQRNKPGVYTRIPEVRDWIKEQTGV"
     variation       360
                     /note="compared with the sequence reported in Acc#
                     AB037898"
                     /replace="t"
     variation       513
                     /note="compared with the sequence reported in Acc#
                     AB037898"
                     /replace="t"
     variation       573
                     /note="compared with the sequence reported in Acc#
                     AB037898"
                     /replace="t"
     variation       2121
                     /note="compared with the sequence reported in Acc#
                     AB037898"
                     /replace="t"
     polyA_site      3157. .3174
BASE COUNT      697 a    887 c    911 g    679 t
ORIGIN

  Query Match             83.5%;  Score 2595;  DB 10;  Length 3174;
  Best Local Similarity   91.1%;  Pred. No. 0;
  Matches 2853;  Conservative    0;  Mismatches  230;  Indels   50;  Gaps    7;

Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
              ||||||| ||| |||| | || ||||||||||||||||||||||||||||| |||| | |
Db         65 CATGGTAGTCGGTTGCCGGCAGAGACCACGCGTCTGAGACCGGCGATCGGGCCGCCGAGA 124

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
              ||||||| | |||||||||||||||||||||||||||| |||||||||||||| ||||
Db        125 CCATGGGGAACAATCGGGGCCGCAAGGCCGGAGGGGGCTCGCAGGACTTCGGCGCCGGAC 184

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 244

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
              ||| ||||| |||||| | ||||||||||||||||||||||||||||| ||||||
Db        245 CTGTGAACAACGCCAAGCAGGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGATGGTGG 304

Qy        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
              | ||  |||||||||| |||||||| |||||||||||||||||||||||||||||||||
Db        305 CCGTAGTGTTCAGCTTCCTTTTGCTCTCGCTCATGGCCGGCTTGCTGGTGTGGCACTTCC 364

Qy        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
              |||| |||||||| || | |||| ||||||||||||||||||||||||| |||||||
Db        365 ATTACCGGAATGTGAGGATCCAAAAGGTCTTCAATGGCCATCTGAGGATCACGAATGAGA 424

Qy        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420
               |||||||| |||||||||||||||||||| ||||| ||||||||||||| ||||
Db        425 ACTTTCTGGACGCGTATGAGAACTCCACCTCCACGGAGTTCATCAGCCTGGCCAGTCAGG 484

Qy        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480
              ||||||| |||||||||| |||||| |||||||||||||||||||||||||||||||||
Db        485 TGAAGGAAGCGCTGAAGCTGATGTACAGTGAAGTCCCTGTCCTGGGCCCCTACCACAAGA 544

Qy        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540
              | ||| |||||||||||||||||||||||| ||||| ||||||||||||||||||||||
Db        545 AATCGACTGTAACTGCCTTCAGTGAGGGCAGCGTCATTGCCTACTACTGGTCAGAGTTCA 604

Qy        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||| |
Db        605 GCATCCCCCCACACCTGGAGGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTCA 664

Qy        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660
              || |||||||||||||||||||||||||||||||||||||| || ||||||||||||||||

Db        665 CACTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTGACCTCTGTGGTGGCCTTCC 724

Qy        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720
              |||||||||||||||||| |||||||||||||||||||||||||||||||||| ||||
Db        725 CCATTGACCCCAGAATGCTCCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTCCATG 784

Qy        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780
              ||| |||     |||||||||| |||||||||||||||||||||| |||||||||||| |
Db        785 CCCGTGGCAGGACAGTGACACGTTTCACTACCCCTGGCTTCCCTAACAGTCCCTACCCAG 844

Qy        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840
              | |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        845 CTCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCGGACTCTGTGCTGAGCCTCACCT 904

Qy        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900
              |||||| |||||||||| |||||||||| ||||| |||||||||||||| |||||||
Db        905 TCCGAAGTTTTGATGTCGCGCCCTGTGATGGGCATGACAGTGACCTGGTCACAGTGTATG 964

Qy        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960
              | |||||||||||||||||| ||||| |||||||||||||||||||||||||||| ||||
Db        965 ACAGCCTGAGCCCCATGGAGCCCCATGCTGTGGTGCGGCTGTGTGGCACCTTCTCGCCCT 1024

Qy        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |
Db       1025 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTTACGCTGATAACCAACA 1084

Qy       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1144

Qy       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140
              |||||||| |||||||||  |||||||| |||||||||||||||||||||||||||| |
Db       1145 GTGGCGGCCTTTTGAGTGAAGCCCAAGGGACGTTTAGCAGCCCCTACTATCCAGGCCATT 1204

Qy       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1205 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAAG 1264

Qy       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
              | |||||||||||||||||||||||||||||||||||| |||||||||  ||||||||||
Db       1265 TACGCTTCAAACTCTTCTATCTGGTGGACCCCAACATACCAGTGGGAAGCTGCACCAAGG 1324

Qy       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1325 ACTATGTGGAGATCAACGGGGAGAAGTTCTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1384

Qy       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1385 GCAACAGCAGCAAGATTACGGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1444

Qy       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
              |||| |||||||| |||||||||||||||||||| ||||| |||||||||||||||| |
Db       1445 TCCTGGCTGAGTATCTCTCCTACGACTCCAATGACCCATGCCCAGGGATGTTCATGTGTA 1504

Qy       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
              |||||||||||||||||| ||||| |||||||||| |||||||| |||||||||| | |
Db       1505 AGACTGGACGGTGCATCCGCAAGGACTTGCGCTGCGATGGCTGGGCTGACTGCCCGGACT 1564

Qy       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
              | |||||||||||| |||||||||||||| || |||||||||||| |||||| |||||||
Db       1565 ACAGTGATGAGCGTCACTGCCGATGCAACGCTACCCACCAGTTCATGTGCAAGAACCAGT 1624

Qy       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
              |||||||||||||||||||||||||||||||||||||||||||||||| || || | |||
Db       1625 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGATGGCAGCGATG 1684

Qy       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
              |||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db       1685 AGGAGGGCTGCAGTTGTCCTGCCGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1744

Qy       1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
              |||||||| |||||||||||||||| |||||||||||||||||||||||||| |||||||
Db       1745 AGAGCCAGCAGTGTAATGGGAAGGACGACTGTGGAGATGGGTCTGACGAGGCCTCATGTG 1804
```

```
Qy      1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
             |||  |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1805 ACAATGTGAATGCCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1864

Qy      1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
             ||| ||||||||||||||||||||||||||||| ||||| |||||||||| |||||||
Db      1865 TGAACAAGGGCAACCCTGAGTGTGATGGGAAGAAGGACTGCAGCGATGGCTCGGATGAGA 1924

Qy      1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      1925 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGAGTGGTTGGTGGCACGA 1984

Qy      1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      1985 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCTCTGGGCCAGGGCCACT 2044

Qy      1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2045 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2104

Qy      2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100
             |||| | || |||||||||||||||| |||| |||||||| |||||||| |||||||||
Db      2105 ATGAGACAATTTTCAAGTACTCAGACCACACCATGTGGACAGCCTTCCTCGGTCTGCTGG 2164

Qy      2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160
             |||||||||||||||||||||||||| |||||||  ||||||||||||||||||||| |
Db      2165 ACCAGAGCAAGCGCAGTGCCTCTGGGGTACAGGAGCACAAGCTCAAACGTATCATCACTC 2224

Qy      2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220
             |||||||||| |||| || |||||||||||||||| ||||||||||||||||||||||
Db      2225 ACCCTTCCTTTAATGACTTTACCTTCGACTATGACATTGCCTTGCTGGAGCTGGAGAAGC 2284

Qy      2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280
             |||  |||||||||| |||||||||||||||||||||||||  |||||||||||| |
Db      2285 CGGCAGAGTACAGCACTGTCGTGCGCCCCATCTGCCTGCCTGATAATACCCATGTCTTTC 2344

Qy      2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
             |||| ||||| |||||||||||||||||||||||| ||||| ||||| ||||| || ||||
Db      2345 CTGCCGGCAAAGCCATCTGGGTCACAGGCTGGGGCCACACGAAAGAAGGAGGAACTGGAG 2404

Qy      2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400
             | |||||||||||||||| |||||||| |||||||||||||||||||||||||||| |||
Db      2405 CACTGATCCTGCAGAAGGGAGAGATCCGGGTCATCAACCAGACCACCTGTGAGGAACTCT 2464

Qy      2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2465 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2524

Qy      2461 CCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
             |||||||||||| || |||||||||||||| || | |||||||||||||||||| ||||
Db      2525 CCTGCCAGGGTGATTCCGGTGGCCCCTTGTCCAGTGTGGAGAAAGATGGGCGAATCTTCC 2584

Qy      2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2585 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2644

Qy      2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
             ||||| |||||| |||||||||||||||||||||||| |||||||||| ||||||||
Db      2645 CAAGGATCCCTGAAGTTCGGGACTGGATCAAAGAGCAAACTGGGGTATAGAAGCATGGAC 2704

Qy      2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
             ||||| || |||| ||| | || |||||||||| || ||  |||||||||| ||||
Db      2705 AGACAACCCACCATGAACTCGCAAAGGGATGCCCCCCACGCGTACCTGGATACAGCAGAG 2764

Qy      2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
             || | |||| ||||||| |||||||||||| |||| |||| ||||||||||||||||||
Db      2765 GATCGCTGAGGACATTTGTGCTGTGGCCTCTCCCC-----CACATCCCAGACTGTGAACT 2819

Qy      2761 GCATCCTTAGGACTCAGAGTTCTTCCA---------AAGTGGGACCCCCTCAAGAGTTGG 2811
             || ||| |||||||||| | |||||          ||  ||  |||||| |||||||
Db      2820 GCTTCCCTAGGACTCAGAATCCTTCCCTGAGCTTCCAAAGTGGGACCCCTCAGGAGTTGG 2879

Qy      2812 AGAGAGAACTTGCGTGCTAGCGGCCCAGCCTG----GGGGCAAGGGTTTGATGGCAGCCT 2867
             |||| |||  ||||| | |||||||| |    |||||||| ||||||||||||||||
Db      2880 CCAGAGATCTTCTGTGCTGGTGGCCCAGCCCGGGGTGGGGCAAGGATTTGATGGCAGCCT 2939

Qy      2868 TCCCCCTCTAGCCCTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCC----- 2922
             | |||||||| |||||| |||||| |||| |||||| ||| ||||| ||||||||
Db      2940 T-CCCCTCTAACCCTGAACTGGGTAAAGACGATGCTTTCCTGGAGAACTGCTTCCAGACT 2998

Qy      2923 ---------AACTGTCATTGAGCTCCCGGGAGCCCTATGGGAGGAGGGGCTCAGGGTCAC 2973
                      |||| || | | ||||| ||| |||||||||  |||||||   ||
Db      2999 CTAAAGGAAAACTATCGTCGCGCTCCTGGGGGCCCTATGGGCAGAGGGGCTGCAGG---- 3054

Qy      2974 TCTTTTCAGGAAGCGCCAGCCCTAGGAACCCCAGAAAAGAGTGGTACCTAAGGCTGAAAT 3033
             |||||||||| |||||||||| |||||||||||||||| ||||| ||||||  |
Db      3055 GTCTTTCAGGAAGCTCCAGCCCTAGGGACCCCAGAAAAGAGTGGCACCTATGGCTGAGTT 3114

Qy      3034 TGTTTTGCTGTTGCCAGGGGTGGGTATTTGAGAGTAAAACATTTTATTTCTTTTTAAAAA 3093
             |||||||||                |||||||||||||||||||||||||||||||||||
Db      3115 TGTTTTGCT-------------GGTATTTGAGAGTAAAACATTTTATTTCTTTTTAAAAA 3161

Qy      3094 AAAAAAAAAAAAA 3106
             |||||||||||||
Db      3162 AAAAAAAAAAAAA 3174
```

```
RESULT 5
AB037898
LOCUS       AB037898               2568 bp    mRNA    linear   ROD 20-OCT-2001
DEFINITION  Rattus norvegicus MBSP mRNA for membrane bound serine protease,
            complete cds.
ACCESSION   AB037898
VERSION     AB037898.1  GI:9650963
KEYWORDS    membrane bound serine protease; MT-SP1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Satomi,S., Yamasaki,Y., Tsuzuki,S., Hitomi,Y., Iwanaga,T. and
            Fushiki,T.
  TITLE     A role for membrane-type serine protease (MT-SP1) in intestinal
            epithelial turnover
  JOURNAL   Biochem. Biophys. Res. Commun. 287 (4), 995-1002 (2001)
  MEDLINE   21458307
   PUBMED   11573963
REFERENCE   2  (bases 1 to 2568)
  AUTHORS   Tsuzuki,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2000) Satoshi Tsuzuki, Kyoto University, Graduate
            School of Agriculture, Division of Food Science and Biotechnology,
            Laboratory of Nutrition Chemistry; Oiwake-cho, Kitashirakawa,
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            (E-mail:tkmone@kais.kais.kyoto-u.ac.jp, Tel:81-75-753-6263,
            Fax:81-75-753-6264)
FEATURES             Location/Qualifiers
     source          1. .2568
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /sex="male"
                     /tissue_type="jejunum"
                     /dev_stage="8 weeks old"
     gene            1. .2568
                     /gene="MBSP"
     CDS             1. .2568
                     /gene="MBSP"
                     /codon_start=1
                     /product="membrane bound serine protease"
                     /protein_id="BAB03502.1"
                     /db_xref="GI:9650964"
                     /translation="MGNNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPVNNA
```

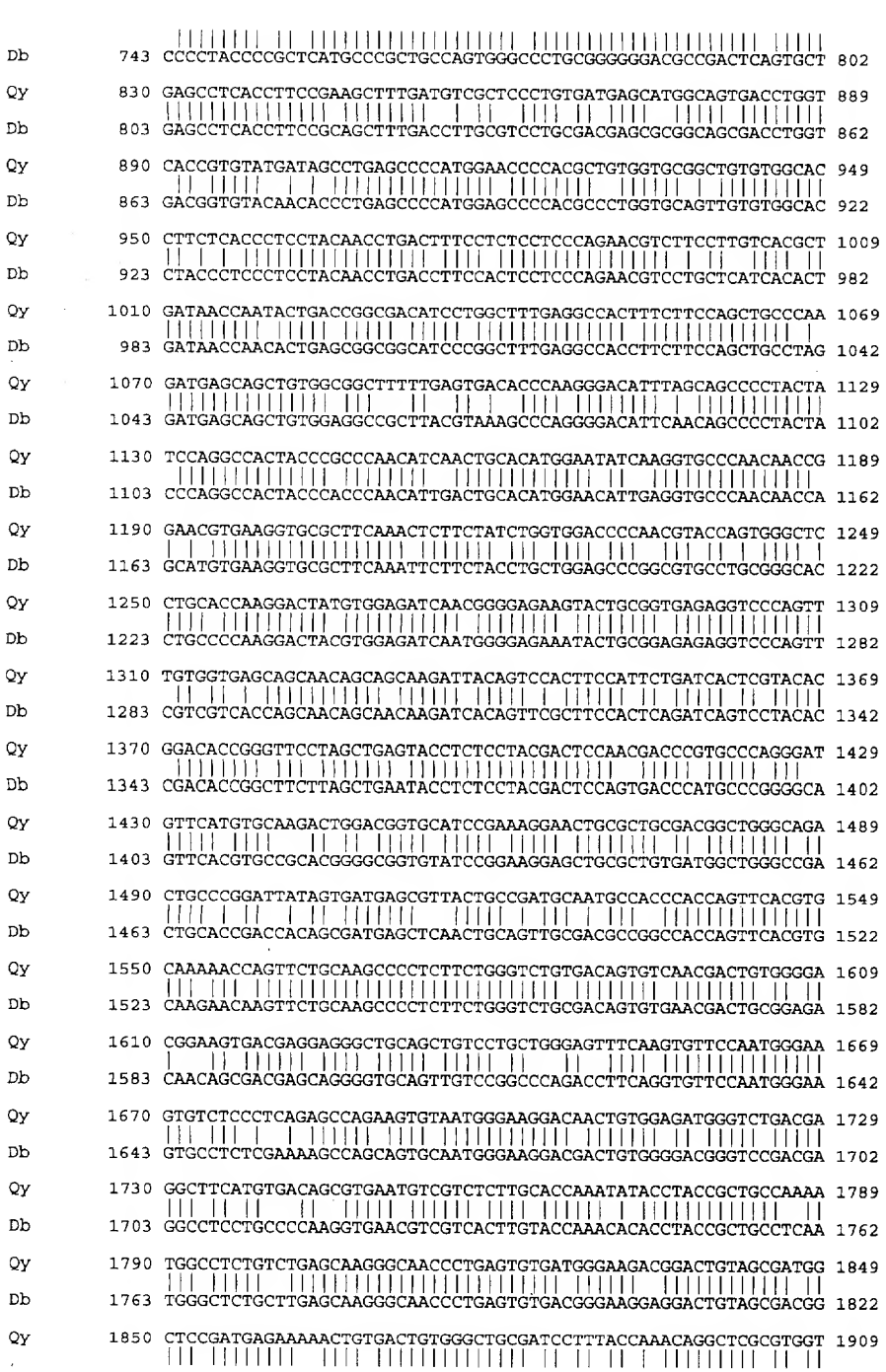
```
                KQVEKRGPRRWVVMVAVVFSFLLLSLMAGLLVWHFHYRNVRIQKVFNGHLRITNENFL
                DAYENSTSTEFISLASQVKEALKLMYSEVPVLGPYHKKSTVTAFSEGSVIAYYWSEFS
                IPPHLEEEVDRAMAVERVVTLPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFAL
                HARGRTVTRFTTPGFPNSPYPAHARCQWVLRGDADSVLSLTFRSFDVAPCDGHDSDLV
                TVYDSLSPMEPHAVVRLCGTFSPSYNLTFLSSQNVFLVTLITNTDRRHPGFEATFFQL
                PKMSSCGGLLSEAQGTFSSPYYPGHYPPNINCTWNIKVPNNRNVKVRFKLFYLVDPNI
                PVGSCTKDYVEINGEKFCGERSQFVVSSNSSKITVHFHSDHSYTDTGFLAEYLSYDSN
                DPCPGMFMCKTGRCIRKDLRCDGWADCPDYSDERHCRCNATHQFMCKNQFCKPLFWVC
                DSVNDCGDGSDEEGCSCPAGSFKCSNGKCLPQSQQCNGKDDCGDGSDEASCDNVNAVS
                CTKYTYRCQNGLCLNKGNPECDGKKDCSDGSDEKNCDCGLRSFTKQARVVGGTNADEG
                EWPWQVSLHALGQGHLCGASLISPDWLVSAAHCFQDETIFKYSDHTMWTAFLGLLDQS
                KRSASGVQEHKLKRIITHPSFNDFTFDYDIALLELEKPAEYSTVVRPICLPDNTHVFP
                AGKAIWVTGWGHTKEGGTGALILQKGEIRVINQTTCEELLPQQITPRMMCVGFLSGGV
                DSCQGDSGGPLSSVEKDGRIFQAGVVSWGEGCAQRNKPGVYTRIPEVRDWIKEQTGV"
BASE COUNT      559 a    708 c    746 g    555 t
ORIGIN

  Query Match            74.5%;  Score 2313.6;  DB 10;  Length 2568;
  Best Local Similarity  93.8%;  Pred. No. 0;
  Matches 2409;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy         63 ATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGACTC 122
Db          1 ATGGGGAACAATCGGGGCCGCAAGGCCGGAGGGGGCTCGCAGGACTTCGGCGCCGGACTC 60

Qy        123 AAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGCCT 182
Db         61 AAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGCCT 120

Qy        183 GCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGGCA 242
Db        121 GTGAACAACGCCAAGCAGGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGATGGTGGCC 180

Qy        243 GTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCCAT 302
Db        181 GTAGTGTTCAGCTTCCTTTTGCTCTCGCTCATGGCCGGCTTGCTGGTGTGGCACTTCCAT 240

Qy        303 TATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGATC 362
Db        241 TACCGGAATGTGAGGATCCAAAAGGTCTTCAATGGCCATCTGAGGATCACGAATGAGAAC 300

Qy        363 TTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGGTG 422
Db        301 TTTCTGGACGCGTATGAGAACTCCACCTCCACGGAGTTCATCAGCCTGGCCAGTCAGGTT 360

Qy        423 AAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGAAG 482
Db        361 AAGGAAGCGCTGAAGCTGATGTACAGTGAAGTCCCTGTCCTGGGCCCCTACCACAAGAAA 420

Qy        483 TCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCAGC 542
Db        421 TCGACTGTAACTGCCTTCAGTGAGGGCAGCGTCATTGCCTACTACTGGTCAGAGTTCAGC 480

Qy        543 ATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAACA 602
Db        481 ATCCCCCCACACCTGGAGGAAGAGGTTGATCGTGCCATGGCTGTGGAGCGAGTTGTCACA 540

Qy        603 TTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCCCC 662
Db        541 CTGCCACCCCGAGCACGGGCACTGAAATCCTTTGTGCTGACCTCTGTGGTGGCCTTCCCC 600

Qy        663 ATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATGCC 722
Db        601 ATTGACCCCAGAATGCTCCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTCCATGCC 660

Qy        723 CATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGGCG 782
Db        661 CGTGGCAGGACAGTGACACGTTTCACTACCCCTGGCTTCCCTAACAGTCCCTACCCAGCT 720

Qy        783 CATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCTTC 842
Db        721 CATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCGGACTCTGTGCTGAGCCTCACCTTC 780

Qy        843 CGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATGAT 902
Db        781 CGAAGTTTTGATGTCGCGCCCTGTGATGGGCATGACAGTGACCTGGTCACAGTGTATGAC 840

Qy        903 AGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCTCC 962
Db        841 AGCCTGAGCCCCATGGAGCCCCATGCTGTGGTGCGGCTGTGTGGCACCTTCTCGCCCTCC 900

Qy        963 TACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATACT 1022
Db        901 TACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTTACGCTGATAACCAACACT 960

Qy       1023 GACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGT 1082
Db        961 GACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGT 1020

Qy       1083 GGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACTAC 1142
Db       1021 GGCGGCCTTTTGAGTGAAGCCCAAGGGACGTTTAGCAGCCCCTACTATCCAGGCCATTAC 1080

Qy       1143 CCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTG 1202
Db       1081 CCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAAGTA 1140

Qy       1203 CGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGGAC 1262
Db       1141 CGCTTCAAACTCTTCTATCTGGTGGACCCCAACATACCAGTGGGAAGCTGCACCAAGGAC 1200

Qy       1263 TATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGC 1322
Db       1201 TATGTGGAGATCAACGGGGAGAAGTTCTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGC 1260

Qy       1323 AACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGTTC 1382
Db       1261 AACAGCAGCAAGATTACGGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGTTC 1320

Qy       1383 CTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAG 1442
Db       1321 CTGGCTGAGTATCTCTCCTACGACTCCAATGACCCATGCCCAGGGATGTTCATGTGTAAG 1380

Qy       1443 ACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTAT 1502
Db       1381 ACTGGACGGTGCATCCGCAAGGACTTGCGCTGCGATGGCTGGGCTGACTGCCCGGACTAC 1440

Qy       1503 AGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGTTC 1562
Db       1441 AGTGATGAGCGTCACTGCCGATGCAACGCTACCCACCAGTTCATGTGCAAGAACCAGTTC 1500

Qy       1563 TGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACGAG 1622
Db       1501 TGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGATGGCAGCGATGAG 1560

Qy       1623 GAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAG 1682
Db       1561 GAGGGCTGCAGTTGTCCTGCCGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAG 1620

Qy       1683 AGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTGAC 1742
Db       1621 AGCCAGCAGTGTAATGGGAAGGACGACTGTGGAGATGGGTCTGACGAGGCCTCATGTGAC 1680

Qy       1743 AGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTG 1802
Db       1681 AATGTGAATGCCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTG 1740

Qy       1803 AGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAA 1862
Db       1741 AACAAGGGCAACCCTGAGTGTGATGGGAAGAAGGACTGCAGCGATGGCTCGGATGAGAAA 1800

Qy       1863 AACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAAT 1922
Db       1801 AACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGAGTGGTTGGTGGCACGAAT 1860

Qy       1923 GCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTG 1982
```

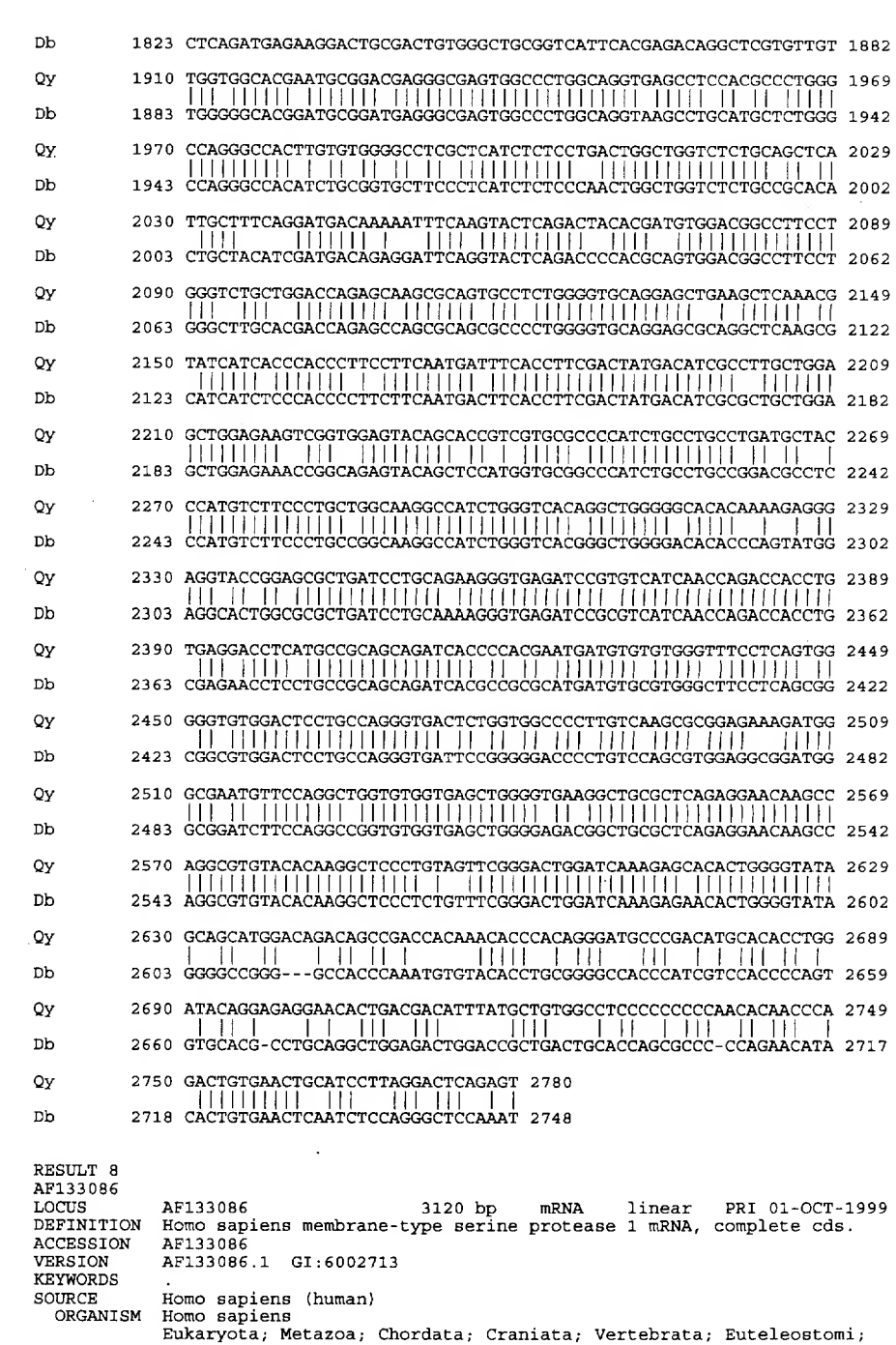
```
Db       1861 GCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCTCTGGGCCAGGGCCACTTG 1920

Qy       1983 TGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGAT 2042

Db       1921 TGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGAT 1980

Qy       2043 GACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGAC 2102

Db       1981 GAGACAATTTTCAAGTACTCAGACCACACCATGTGGACAGCCTTCCTCGGTCTGCTGGAC 2040

Qy       2103 CAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCCAC 2162

Db       2041 CAGAGCAAGCGCAGTGCCTCTGGGGTACAGGAGCACAAGCTCAAACGTATCATCACTCAC 2100

Qy       2163 CCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGTCG 2222

Db       2101 CCTTCCTTTAATGACTTTACTTTCGACTATGACATTGCCTTGCTGGAGCTGGAGAAGCCG 2160

Qy       2223 GTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCCCT 2282

Db       2161 GCAGAGTACAGCACTGTCGTGCGCCCCATCTGCCTGCCTGATAATACCCATGTCTTTCCT 2220

Qy       2283 GCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAGCG 2342

Db       2221 GCCGGCAAAGCCATCTGGGTCACAGGCTGGGGCCACACGAAAGAAGGAGGAACTGGAGCA 2280

Qy       2343 CTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCATG 2402

Db       2281 CTGATCCTGCAGAAGGGAGAGATCCGGGTCATCAACCAGACCACCTGTGAGGAACTCTTG 2340

Qy       2403 CCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCC 2462

Db       2341 CCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCC 2400

Qy       2463 TGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAG 2522

Db       2401 TGCCAGGGTGATTCCGGTGGCCCCTTGTCCAGTGTGGAGAAAGATGGGCGAATCTTCCAG 2460

Qy       2523 GCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACACA 2582

Db       2461 GCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACACA 2520

Qy       2583 AGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAG 2630

Db       2521 AGGATCCCTGAAGTTCGGGACTGGATCAAAGAGCAAACTGGGGTATAG 2568
```

```
RESULT 6
BC030532
LOCUS       BC030532               3273 bp    mRNA    linear   PRI 20-MAY-2002
DEFINITION  Homo sapiens, suppression of tumorigenicity 14 (colon carcinoma,
            matriptase, epithin), clone MGC:40392 IMAGE:5213189, mRNA, complete
            cds.
ACCESSION   BC030532
VERSION     BC030532.1  GI:20988874
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3273)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Laric,P., Legaspi,R., Maduro,Q.L.,
            Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C., McDowell,J.,
            Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C.,
            Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A.,
            Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 64 Row: k Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .3273
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:6768"
                     /db_xref="taxon:9606"
                     /clone="MGC:40392 IMAGE:5213189"
                     /tissue_type="Blood, adult leukocytes"
                     /clone_lib="NIH_MGC_118"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             118. .2685
                     /codon_start=1
                     /product="suppression of tumorigenicity 14 (colon
                     carcinoma, matriptase, epithin)"
                     /protein_id="AAH30532.1"
                     /db_xref="GI:20988875"
                     /translation="MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNV
                     KKVEKHGPGRWVVLAAVLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFV
                     DAYENSNSTEFVSLASKVKDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFS
                     IPQHLVEEAERVMAEERVVMLPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGL
                     HARGVELMRFTTPGFPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLV
                     TVYNTLSPMEPHALVQLCGTYPPSYNLTFHSSQNVLLITLITNTERRHPGFEATFFQL
                     PRMSSCGGRLRKAQGTFNSPYYPGHYPPNIDCTWNIEVPNNQHVKVRFKFFYLLEPGV
                     PAGTCPKDYVEINGEKYCGERSQFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSS
                     DPCPGQFTCRTGRCIRKELRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVC
                     DSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVVT
                     CTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQARVVGGTDADEG
                     EWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFLGLHDQS
                     QRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFP
                     AGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFLSGGV
                     DSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTGV"
BASE COUNT      690 a   1001 c    993 g    589 t
ORIGIN
```

```
  Query Match           61.4%;  Score 1905.6;  DB 9;  Length 3273;
  Best Local Similarity  81.3%;  Pred. No. 0;
  Matches 2260;  Conservative    0;  Mismatches  514;  Indels    6;  Gaps    4;

Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60

Db         57 CATGGGAGCCGGCCGCCGGCAGGGACGACGCCTGTGAGACCCGCGAGCGG-CCTCGGGGA 115

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120

Db        116 CCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAGGACTTCGGCGCGGGAC 175

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180

Db        176 TCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAAGGCGTGGAGTTCCTGC 235

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
```

```
Qy       2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
             ||||||||||||||||| || || |||||||| ||||| |||||||| || || |||||||
Db       2456 TGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTCAGCGGCGGCGTGGACT 2515

Qy       2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
             |||||||||||||| || || || ||| |||| |||| ||||   |||||||| || ||||
Db       2516 CCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCGGATGGGCGGATCTTCC 2575

Qy       2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
             |||| ||||||||||||||||| || ||||||||||||||||||||||||||||||||||||
Db       2576 AGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2635

Qy       2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
             |||||||||||  |||||||||||||||||||| ||||||||||||||   ||   ||
Db       2636 CAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGGGTATAGGGGCCGGG-- 2693

Qy       2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
              | || || |      |||||  | |||   |||  | | ||| || |  | || |
Db       2694 -GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACCCCAGTGTGCACG-CCT 2751

Qy       2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
              | |  |||  |||      ||||    | ||  | |||  || |||  | ||||||||||
Db       2752 GCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGAACATACACTGTGAACT 2810

Qy       2761 GCATCCTTAGGACTCAGAGT 2780
                |||   ||| |||  | |
Db       2811 CAATCTCCAGGGCTCCAAAT 2830


RESULT 7
AF118224
LOCUS       AF118224                3149 bp    mRNA    linear   PRI 30-DEC-1999
DEFINITION  Homo sapiens matriptase mRNA, complete cds.
ACCESSION   AF118224
VERSION     AF118224.2  GI:6647301
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2955)
  AUTHORS   Lin,C.Y., Anders,J., Johnson,M., Sang,Q.A. and Dickson,R.B.
  TITLE     Molecular cloning of cDNA for matriptase, a matrix-degrading serine
            protease with trypsin-like activity
  JOURNAL   J. Biol. Chem. 274 (26), 18231-18236 (1999)
  MEDLINE   99303581
   PUBMED   10373424
REFERENCE   2  (bases 1 to 2955)
  AUTHORS   Lin,C.Y., Anders,J., Johnson,M., Sang,Q.A. and Dickson,R.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-1999) Lombardi Cancer Center, Georgetown
            University Medical Center, 3970 Reservoir Road NW, Washington, DC
            20007, USA
REFERENCE   3  (bases 1 to 3149)
  AUTHORS   Oberst,M.D., Dickson,R.B. and Lin,C.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Lombardi Cancer Center, Georgetown
            University Medical Center, 3970 Reservoir Road NW, Washington, DC
            20007, USA
  REMARK    Sequence update by submitter
COMMENT     On Dec 30, 1999 this sequence version replaced gi:5359674.
FEATURES             Location/Qualifiers
     source          1. .3149
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             36. .2603
                     /function="degrade extracellular matrix"
                     /note="serine protease with trypsin-like activity;
                     contains C1r/s and LDL receptor domains"
                     /codon_start=1
                     /product="matriptase"
                     /protein_id="AAD42765.2"
                     /db_xref="GI:6647302"
                     /translation="MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNV
                     KKVEKHGPGRWVVLAAVLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFV
                     DAYENSNSTEFVSLASKVKDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFS
                     IPQHLVEEAERVMAEERVVMLPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGL
                     HARGVELMRFTTPGFPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLV
                     TVYNTLSPMEPHALVQLCGTYPPSYNLTFHSSQNVLLITLITNTERRHPGFEATFFQL
                     PRMSSCGGRLRKAQGTFNSPYYPGHYPPNIDCTWNIEVPNNQHVKVRFKFFYLLEPGV
                     PAGTCPKDYVEINGEKYCGERSQFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSS
                     DPCPGQFTCRTGRCIRKELRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVC
                     DSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVVT
                     CTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQARVVGGTDADEG
                     EWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFLGLHDQS
                     QRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFP
                     AGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFLSGGV
                     DSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTGV"
BASE COUNT      647 a    959 c    962 g    581 t
ORIGIN

  Query Match           60.8%;  Score 1889.8;  DB 9;  Length 3149;
  Best Local Similarity  81.4%;  Pred. No. 0;
  Matches 2222;  Conservative    0;  Mismatches  504;  Indels    5;  Gaps    3;

Qy         50 GACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTT 109
              | ||  |    |||||||| ||| |||||| |||||||| |||||||||  |  |||||||
Db         23 GGCCTCGGGGACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAGGACTT 82

Qy        110 CGGCGCGGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGT 169
              |||||||||||||||||||||||||||||  |||||  |||||||||  |||||  ||  ||
Db         83 CGGCGCGGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAAGGCGT 142

Qy        170 GGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGT 229
              ||||||||||| |  |||||  | ||||||  |||||  ||||   |||||  ||||||||||
Db        143 GGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGCTGGGT 202

Qy        230 GGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGT 289
              ||||||||  ||  ||||||  ||  ||  ||||||||  |||   ||    |   |||||  |||||
Db        203 GGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTCCTGGT 262

Qy        290 GTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGAT 349
              ||||||  ||  ||  ||  |||  | |||||  ||  ||  ||  ||||||||||| |  |||||||
Db        263 GTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATGAGGAT 322

Qy        350 CACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCT 409
              ||||||||||  |||  |||||||  ||  ||||||||||  ||||||  ||||||  | |||||
Db        323 CACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTAAGCCT 382

Qy        410 GGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCC 469
              |||||||  |||||||||  ||||||||||||||||  |  ||||||  |||||||  ||
Db        383 GGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTGGGCCC 442

Qy        470 CTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTG 529
              ||||||||||  ||||||||||  ||  ||||||||  ||||||||  ||||||||||||||||||
Db        443 CTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTACTACTG 502

Qy        530 GTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGA 589
              |||  ||||||||||||||  |  |||||||  ||  ||||  ||  ||||  ||||||  | |||
Db        503 GTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCCGAGGA 562

Qy        590 GCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGT 649
              |||  ||  ||  |   ||||  ||  ||  ||  ||  |  |||||  |||||  |||  |  ||  ||
Db        563 GCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACCTCAGT 622

Qy        650 GGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTT 709
              ||||||  |||||||   |||  |||  ||    |  ||||||||  ||||||||||||||||||  ||
Db        623 GGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGCAGCTT 682

Qy        710 TGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAG 769
              ||  ||||||  ||||   ||||  |   |||  ||||||||  ||  ||  ||||||||   ||||
Db        683 TGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCTGACAG 742

Qy        770 TCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCT 829
```



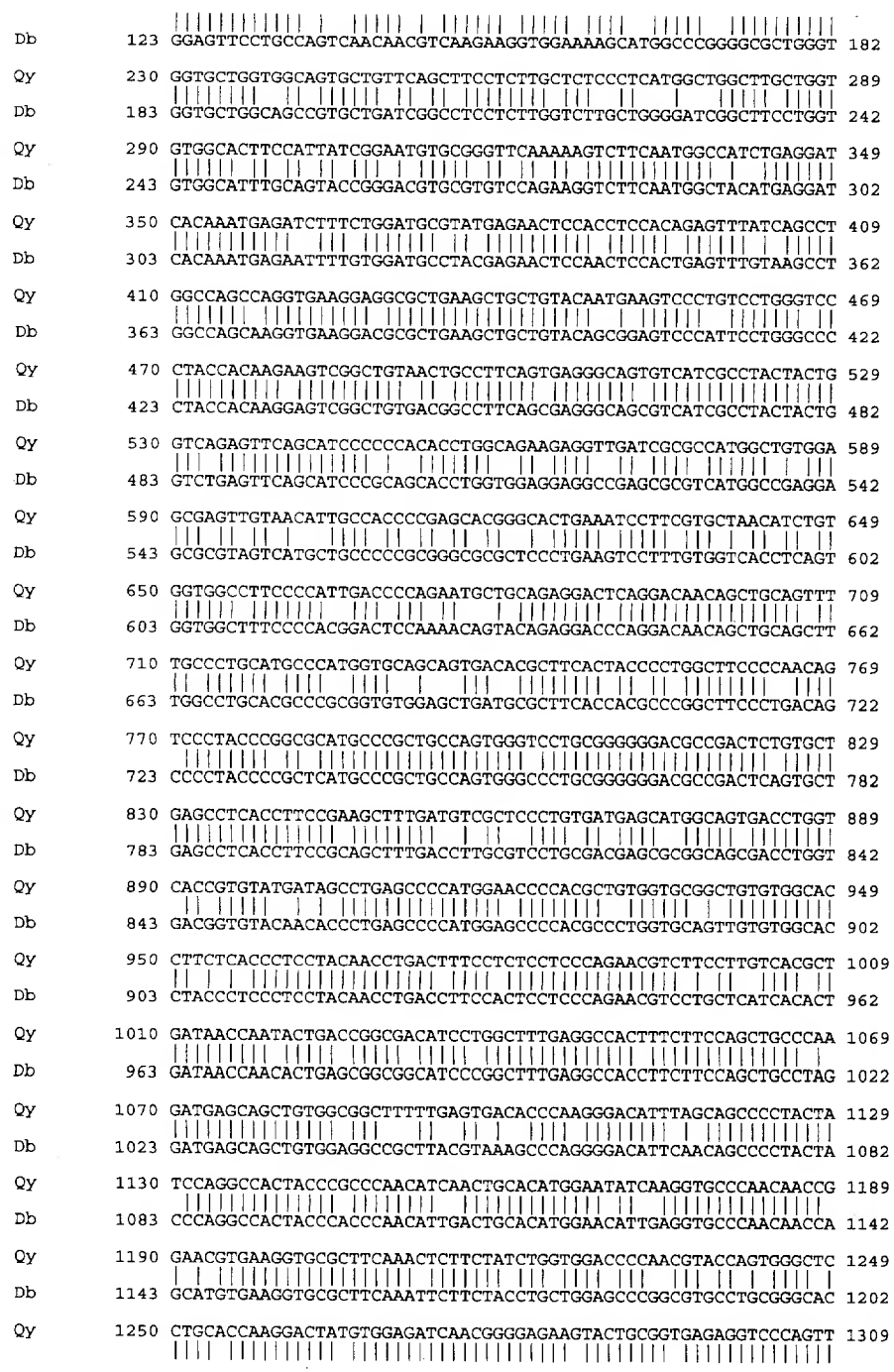


RESULT 8
AF133086
LOCUS AF133086 3120 bp mRNA linear PRI 01-OCT-1999
DEFINITION Homo sapiens membrane-type serine protease 1 mRNA, complete cds.
ACCESSION AF133086
VERSION AF133086.1 GI:6002713
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

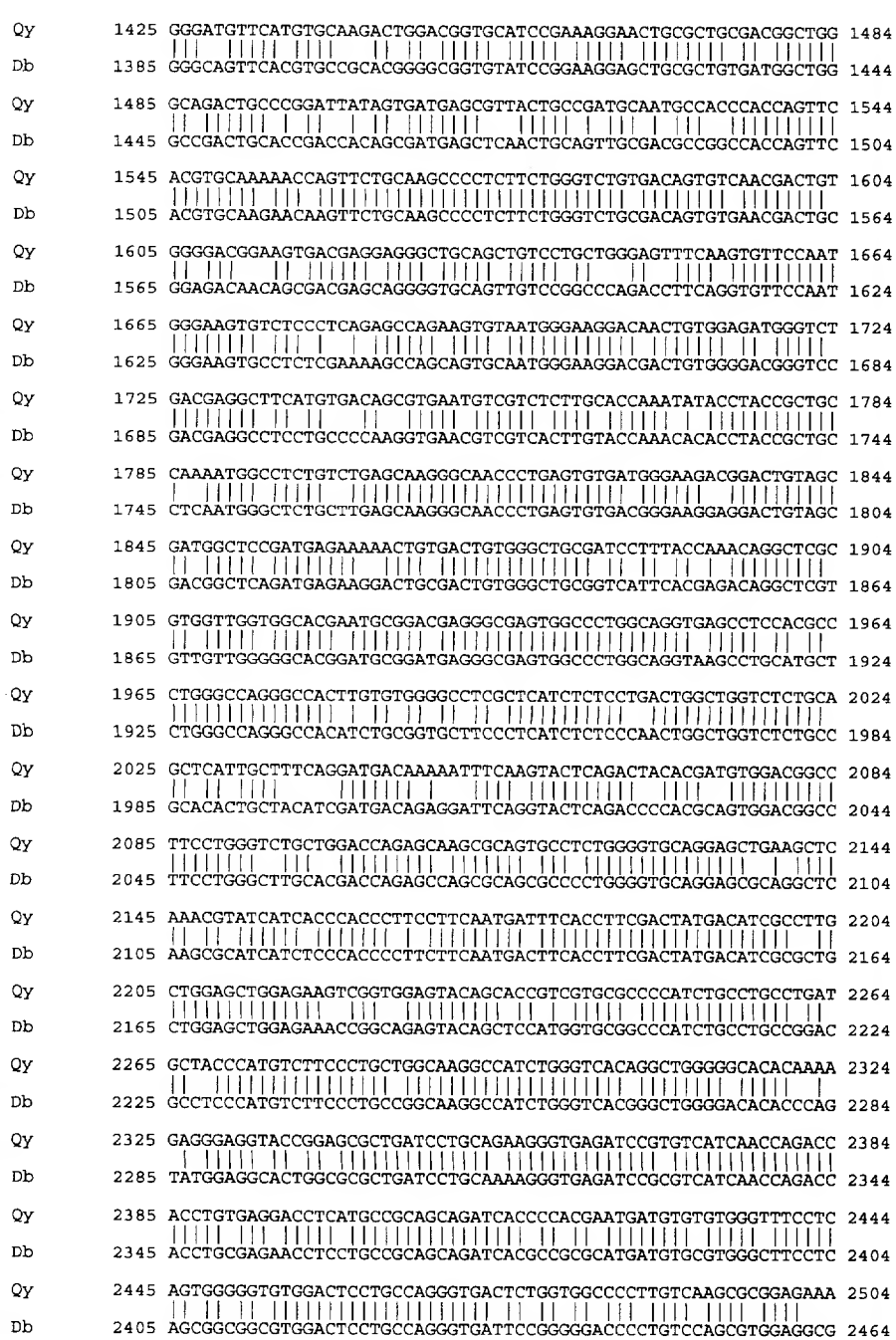
```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3120)
  AUTHORS   Takeuchi,T., Shuman,M.A. and Craik,C.S.
  TITLE     Reverse biochemistry: use of macromolecular protease inhibitors to
            dissect complex biological processes and identify a membrane-type
            serine protease in epithelial cancer and normal tissue
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11054-11061 (1999)
  MEDLINE   99432178
   PUBMED   10500122
REFERENCE   2  (bases 1 to 3120)
  AUTHORS   Takeuchi,T., Shuman,M.A. and Craik,C.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-1999) Dept. Pharm. Chem., University of
            California, San Francisco, 513 Parnassus Ave., Box 0446, San
            Francisco, CA 94143-0446, USA
FEATURES             Location/Qualifiers
     source          1. .3120
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate carcinoma"
     CDS             37. .2604
                     /note="MT-SP1; chymotrypsin clan serine protease"
                     /codon_start=1
                     /product="membrane-type serine protease 1"
                     /protein_id="AAF00109.1"
                     /db_xref="GI:6002714"
                     /translation="MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNV
                     KKVEKHGPGRWVVLAAVLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFV
                     DAYENSNSTEFVSLASKVKDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFS
                     IPQHLVEEAERVMAEERVVMLPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGL
                     HARGVELMRFTTPGFPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLV
                     TVYNTLSPMEPHALVQLCGTYPPSYNLTFHSSQNVLLITLITNTERRHPGFEATFFQL
                     PRMSSCGGRLRKAQGTFNSPYYPGHYPPNIDCTWNIEVPNNQHVKVRFKFFYLLEPGV
                     PAGTCPKDYVEINGEKYCGERSQFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSS
                     DPCPGQFTCRTGRCIRKELRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVC
                     DSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVVT
                     CTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQARVVGGTDADEG
                     EWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFLGLHDQS
                     QRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFP
                     AGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFLSGGV
                     DSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTGV"
BASE COUNT      629 a    956 c    964 g    571 t
ORIGIN

  Query Match           60.7%;  Score 1884.8;  DB 9;  Length 3120;
  Best Local Similarity  82.8%;  Pred. No. 0;
  Matches 2164;  Conservative    0;  Mismatches  447;  Indels    1;  Gaps    1;

Qy         34 CTGAGACCGGCGATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAG 93
Db          9 CTGAGACCCGCGAGCGG-CCTCGGGGACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAG 67

Qy         94 GGGGCTCTCAGGACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATG 153
Db         68 GGGGCCCGAAGGACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATG 127

Qy        154 GCTTTGAGGAGGGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAG 213
Db        128 GCTTGGAGGAAGGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATG 187

Qy        214 GCCCCAGGCGCTGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCA 273
Db        188 GCCCGGGGCGCTGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGG 247

Qy        274 TGGCTGGCTTGCTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCA 333
Db        248 GGATCGGCTTCCTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCA 307

Qy        334 ATGGCCATCTGAGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCA 393
Db        308 ATGGCTACATGAGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCA 367

Qy        394 CAGAGTTTATCAGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAG 453
Db        368 CTGAGTTTGTAAGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAG 427

Qy        454 TCCCTGTCCTGGGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTG 513
Db        428 TCCCATTCCTGGGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCG 487

Qy        514 TCATCGCCTACTACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATC 573
Db        488 TCATCGCCTACTACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGC 547

Qy        574 GCGCCATGGCTGTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCT 633
Db        548 GCGTCATGGCCGAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCT 607

Qy        634 TCGTGCTAACATCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGG 693
Db        608 TTGTGGTCACCTCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGG 667

Qy        694 ACAACAGCTGCAGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCC 753
Db        668 ACAACAGCTGCAGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGC 727

Qy        754 CTGGCTTCCCCAACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGG 813
Db        728 CCGGCTTCCCTGACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGG 787

Qy        814 ACGCCGACTCTGTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGC 873
Db        788 ACGCCGACTCAGTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGC 847

Qy        874 ATGGCAGTGACCTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGG 933
Db        848 GCGGCAGCGACCTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGG 907

Qy        934 TGCGGCTGTGTGGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACG 993
Db        908 TGCAGTTGTGTGGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACG 967

Qy        994 TCTTCCTTGTCACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTT 1053
Db        968 TCCTGCTCATCACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCT 1027

Qy       1054 TCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACAT 1113
Db       1028 TCTTCCAGCTGCCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACAT 1087

Qy       1114 TTAGCAGCCCCTACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCA 1173
Db       1088 TCAACAGCCCCTACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTG 1147

Qy       1174 AGGTGCCCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCA 1233
Db       1148 AGGTGCCCAACAACCAGCATGTGAAGGTGCGCTTCAAATTCTTCTACCTGCTGGAGCCCG 1207

Qy       1234 ACGTACCAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCG 1293
Db       1208 GCGTGCCTGCGGGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCG 1267

Qy       1294 GTGAGAGGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATT 1353
Db       1268 GAGAGAGGTCCCAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACT 1327

Qy       1354 CTGATCACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACG 1413
Db       1328 CAGATCAGTCCTACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTG 1387

Qy       1414 ACCCGTGCCCAGGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCT 1473
Db       1388 ACCCATGCCCGGGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCT 1447
```

```
Qy       1474 GCGACGGCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCA 1533
Db       1448 GTGATGGCTGGGCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCG 1507

Qy       1534 CCCACCAGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTG 1593
Db       1508 GCCACCAGTTCACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTG 1567

Qy       1594 TCAACGACTGTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCA 1653
Db       1568 TGAACGACTGCGGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCA 1627

Qy       1654 AGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTG 1713
Db       1628 GGTGTTCCAATGGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTG 1687

Qy       1714 GAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATA 1773
Db       1688 GGGACGGGTCCGACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACA 1747

Qy       1774 CCTACCGCTGCCAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGA 1833
Db       1748 CCTACCGCTGCCTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGG 1807

Qy       1834 CGGACTGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCA 1893
Db       1808 AGGACTGTAGCGACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGA 1867

Qy       1894 AACAGGCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGA 1953
Db       1868 GACAGGCTCGTGTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAA 1927

Qy       1954 GCCTCCACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGC 2013
Db       1928 GCCTGCATGCTCTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGC 1987

Qy       2014 TGGTCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGA 2073
Db       1988 TGGTCTCTGCCGCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGC 2047

Qy       2074 TGTGGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGG 2133
Db       2048 AGTGGACGGCCTTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGG 2107

Qy       2134 AGCTGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATG 2193
Db       2108 AGCGCAGGCTCAAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATG 2167

Qy       2194 ACATCGCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCT 2253
Db       2168 ACATCGCGCTGCTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCT 2227

Qy       2254 GCCTGCCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGG 2313
Db       2228 GCCTGCCGGACGCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGG 2287

Qy       2314 GGCACACAAAAGAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCA 2373
Db       2288 GACACACCCAGTATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCA 2347

Qy       2374 TCAACCAGACCACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTG 2433
Db       2348 TCAACCAGACCACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCG 2407

Qy       2434 TGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAA 2493
Db       2408 TGGGCTTCCTCAGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCA 2467

Qy       2494 GCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCG 2553
Db       2468 GCGTGGAGGCGGATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCG 2527

Qy       2554 CTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAG 2613
Db       2528 CTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAG 2587

Qy       2614 AGCACACTGGGGTATAGCAGCATGGACAGACA 2645
Db       2588 AGAACACTGGGGTATAGGGGCCGGGGCACCCA 2619
```

```
RESULT 9
AB030036
LOCUS       AB030036               3128 bp    mRNA    linear   PRI 06-JAN-2001
DEFINITION  Homo sapiens mRNA for prostamin, complete cds.
ACCESSION   AB030036
VERSION     AB030036.1  GI:12249014
KEYWORDS    prostamin.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamaguchi,N. and Mitsui,S.
  TITLE     Molecular cloning of a novel transmembrane serine protease
            expressed in human prostate
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3128)
  AUTHORS   Yamaguchi,N. and Mitsui,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1999) Nozomi Yamaguchi, Kyoto Prefectural
            University of Medicine, Res. Ins. Geriatrics; Kawaramachi Hirokoji,
            Kyoto, Kyoto 602-8566, Japan (E-mail:nozomi@koto.kpu-m.ac.jp,
            Tel:81-75-251-5848, Fax:81-75-251-5848)
FEATURES             Location/Qualifiers
     source          1. .3128
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="prostate"
     CDS             16. .2583
                     /codon_start=1
                     /product="prostamin"
                     /protein_id="BAB20376.1"
                     /db_xref="GI:12249015"
                     /translation="MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNV
                     KKVEKHGPGRWVVLAAVLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFV
                     DAYENSNSTEFVSLASKVKDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFS
                     IPQHLVEEAERVMAEERVVMLPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGL
                     HARGVELMRFTTPGFPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLV
                     TVYNTLSPMEPHALVQLCGTYPPSYNLTFHSSQNVLLITLITNTERRHPGFEATFFQL
                     PRMSSCGGRLRKAQGTFNSPYYPGHYPPNIDCTWNIEVPNNQHVKVRFKFFYLLEPGV
                     PAGTCPKDYVEINGEKYCGERSQFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSS
                     DPCPGQFTCRTGRCIRKELRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVC
                     DSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVVT
                     CTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQARVVGGTDADEG
                     EWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTVFLGLHDQS
                     QRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFP
                     AGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFLSGGV
                     DSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTGV"
BASE COUNT      637 a    954 c    956 g    581 t
ORIGIN

  Query Match           60.7%;  Score 1884.6;  DB 9;  Length 3128;
  Best Local Similarity   81.4%;  Pred. No. 0;
  Matches 2222;  Conservative    0;  Mismatches  504;  Indels    5;  Gaps    3;

Qy          50 GACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTT 109
Db           3 GGCCTCGGGGACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAGGACTT 62

Qy         110 CGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGT 169
Db          63 CGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAAGGCGT 122

Qy         170 GGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGT 229
```

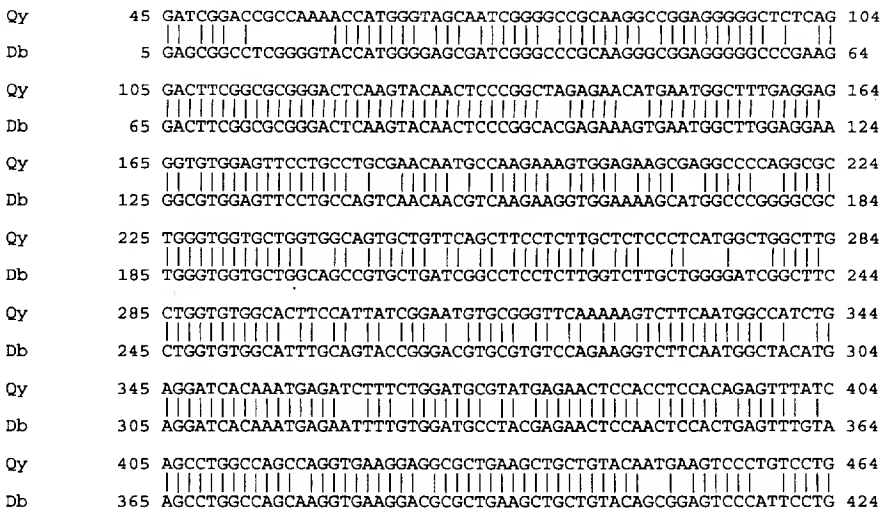

Db 123 GGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGCTGGGT 182
Qy 230 GGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGT 289
Db 183 GGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTCCTGGT 242
Qy 290 GTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGAT 349
Db 243 GTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATGAGGAT 302
Qy 350 CACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCT 409
Db 303 CACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTAAGCCT 362
Qy 410 GGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCC 469
Db 363 GGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTGGGCCC 422
Qy 470 CTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTG 529
Db 423 CTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTACTACTG 482
Qy 530 GTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGA 589
Db 483 GTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCCGAGGA 542
Qy 590 GCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGT 649
Db 543 GCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACCTCAGT 602
Qy 650 GGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTT 709
Db 603 GGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGCAGCTT 662
Qy 710 TGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAG 769
Db 663 TGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCTGACAG 722
Qy 770 TCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCT 829
Db 723 CCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCAGTGCT 782
Qy 830 GAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGT 889
Db 783 GAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGACCTGGT 842
Qy 890 CACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCAC 949
Db 843 GACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGTGGCAC 902
Qy 950 CTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCT 1009
Db 903 CTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATCACACT 962
Qy 1010 GATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAA 1069
Db 963 GATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTGCCTAG 1022
Qy 1070 GATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTA 1129
Db 1023 GATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCCTACTA 1082
Qy 1130 TCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCG 1189
Db 1083 CCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAACAACCA 1142
Qy 1190 GAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTC 1249
Db 1143 GCATGTGAAGGTGCGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCGGGCAC 1202
Qy 1250 CTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTT 1309
Db 1203 CTGCCCCAAGGACTACGTGGAGATCAACGGGGAGAAATACTGCGGAGAGAGGTCCCAGTT 1262
Qy 1310 TGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACAC 1369
Db 1263 CGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCCTACAC 1322
Qy 1370 GGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGAT 1429
Db 1323 CGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCGGGGCA 1382
Qy 1430 GTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGA 1489
Db 1383 GTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGGGCCGA 1442
Qy 1490 CTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTG 1549
Db 1443 CTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTCACGTG 1502
Qy 1550 CAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGA 1609
Db 1503 CAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGCGGAGA 1562
Qy 1610 CGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAA 1669
Db 1563 CAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAATGGGAA 1622
Qy 1670 GTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGA 1729
Db 1623 GTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCCGACGA 1682
Qy 1730 GGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAA 1789
Db 1683 GGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGCCTCAA 1742
Qy 1790 TGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGG 1849
Db 1743 TGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGCGACGG 1802
Qy 1850 CTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGT 1909
Db 1803 CTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGTGTTGT 1862
Qy 1910 TGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGG 1969
Db 1863 TGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCTCTGGG 1922
Qy 1970 CCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCA 2029
Db 1923 CCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCCGCACA 1982
Qy 2030 TTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCT 2089
Db 1983 CTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGTCTTCCT 2042
Qy 2090 GGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACG 2149
Db 2043 GGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTCAAGCG 2102
Qy 2150 TATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGA 2209
Db 2103 CATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTGCTGGA 2162
Qy 2210 GCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTAC 2269
Db 2163 GCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGACGCCTC 2222
Qy 2270 CCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGG 2329
Db 2223 CCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAGTATGG 2282
Qy 2330 AGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTG 2389
Db 2283 AGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACCACCTG 2342

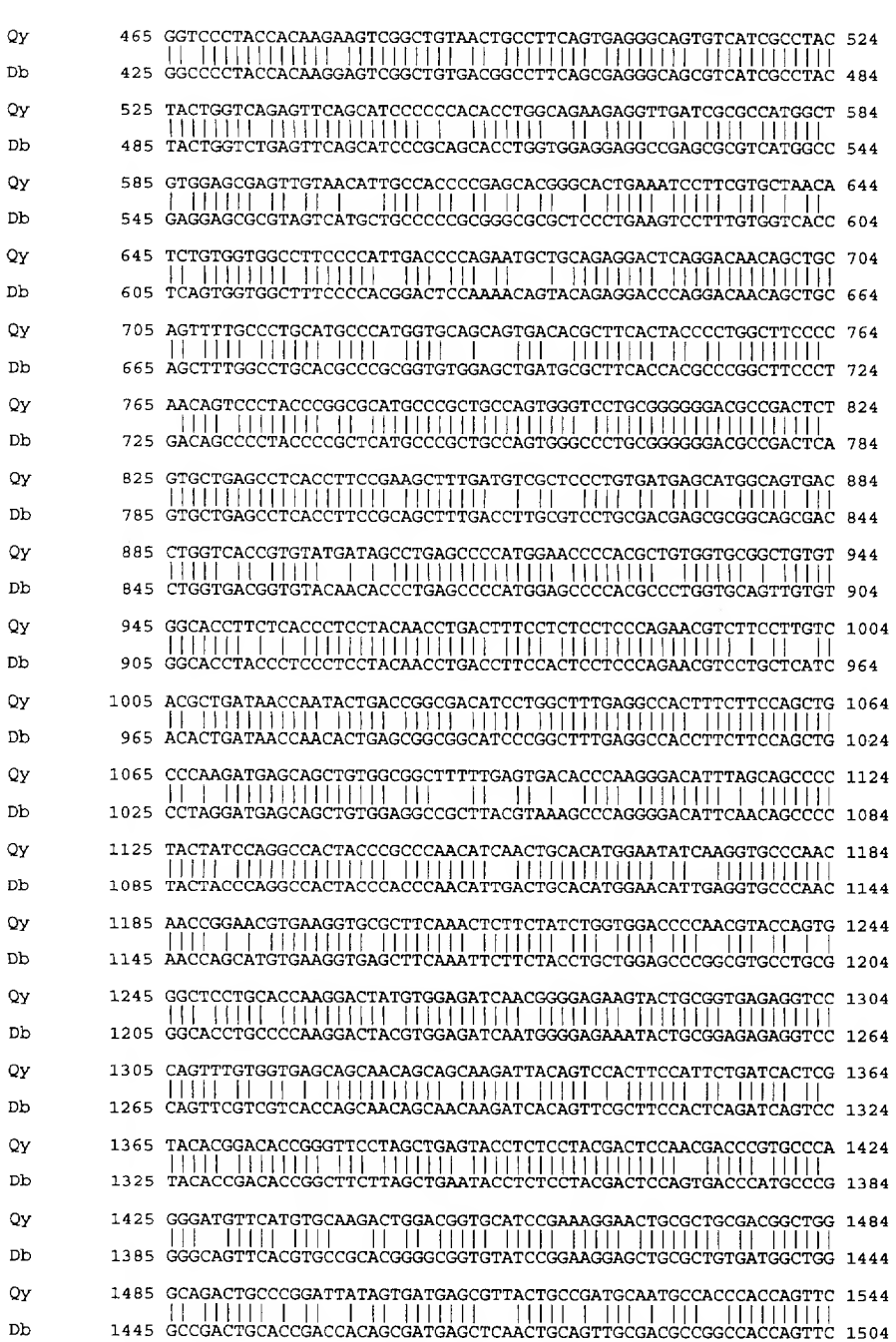
```
Qy       2390 TGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGG 2449
Db       2343 CGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTCAGCGG 2402

Qy       2450 GGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGG 2509
Db       2403 CGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCGGATGG 2462

Qy       2510 GCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCC 2569
Db       2463 GCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAACAAGCC 2522

Qy       2570 AGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATA 2629
Db       2523 AGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGGGTATA 2582

Qy       2630 GCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGG 2689
Db       2583 GGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACCCCAGT 2639

Qy       2690 ATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCA 2749
Db       2640 GTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGAACATA 2697

Qy       2750 GACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2698 CACTGTGAACTCAATCTCCAGGGCTCCAAAT 2728


RESULT 10
AR081724
LOCUS       AR081724               3147 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 1 from patent US 5972616.
ACCESSION   AR081724
VERSION     AR081724.1  GI:10008450
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3147)
  AUTHORS   O'Brien,T.J. and Tanimoto,H.
  TITLE     TADG-15: an extracellular serine protease overexpressed in breast
            and ovarian carcinomas
  JOURNAL   Patent: US 5972616-A 1 26-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .3147
                     /organism="unknown"
BASE COUNT      654 a    952 c    958 g    583 t
ORIGIN

  Query Match           60.6%;  Score 1883.2;  DB 6;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCT 824
Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
```

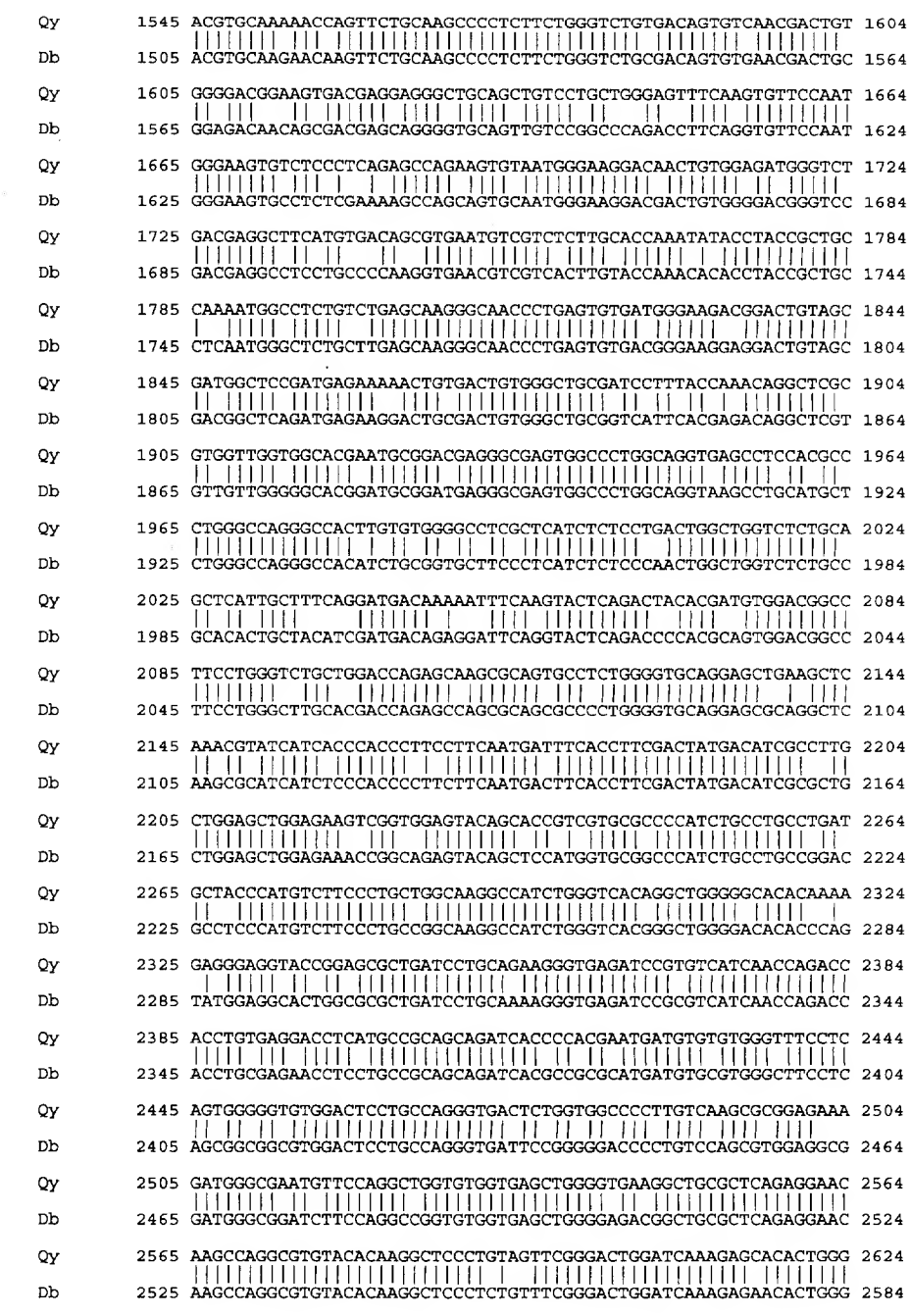


RESULT 11
AR229704
LOCUS AR229704 3147 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 6451500.
ACCESSION AR229704
VERSION AR229704.1 GI:27269434
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 3147)
AUTHORS Leon,J.W.
TITLE Imaging member containing heat switchable carboxylate polymer and method of use
JOURNAL Patent: US 6451500-A 1 17-SEP-2002;
FEATURES Location/Qualifiers
source 1. .3147
/organism="unknown"
BASE COUNT 654 a 952 c 958 g 583 t
ORIGIN

Query Match 60.6%; Score 1883.2; DB 6; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;







```
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 12
AR229712/c
LOCUS       AR229712                3147 bp    RNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 18 from patent US 6451500.
ACCESSION   AR229712
VERSION     AR229712.1  GI:27269442
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3147)
  AUTHORS   Leon,J.W.
  TITLE     Imaging member containing heat switchable carboxylate polymer and
            method of use
  JOURNAL   Patent: US 6451500-A 18 17-SEP-2002;
FEATURES             Location/Qualifiers
     source          1. .3147
                     /organism="unknown"
BASE COUNT      583 a    958 c    952 g    654 t
ORIGIN

  Query Match             60.6%;  Score 1883.2;  DB 6;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104

Db       3143 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 3084

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164

Db       3083 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 3024

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224

Db       3023 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 2964

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284

Db       2963 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 2904

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db       2903 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 2844

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404

Db       2843 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 2784

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464

Db       2783 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 2724

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524

Db       2723 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 2664

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584

Db       2663 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 2604

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644

Db       2603 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 2544

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704

Db       2543 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 2484

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764

Db       2483 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 2424

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824

Db       2423 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 2364

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884

Db       2363 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 2304

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944

Db       2303 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 2244

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004

Db       2243 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 2184

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064

Db       2183 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 2124

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124

Db       2123 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 2064

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184

Db       2063 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 2004

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244

Db       2003 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1944

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304

Db       1943 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1884

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364

Db       1883 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1824

Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424

Db       1823 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1764

Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484

Db       1763 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1704

Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544

Db       1703 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1644

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604

Db       1643 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1584

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664

Db       1583 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1524

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
```

```
Db      1523 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1464

Qy      1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784

Db      1463 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1404

Qy      1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844

Db      1403 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1344

Qy      1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904

Db      1343 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1284

Qy      1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964

Db      1283 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1224

Qy      1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024

Db      1223 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1164

Qy      2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084

Db      1163 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 1104

Qy      2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144

Db      1103 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 1044

Qy      2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204

Db      1043 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 984

Qy      2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264

Db       983 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 924

Qy      2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324

Db       923 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 864

Qy      2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384

Db       863 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 804

Qy      2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444

Db       803 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 744

Qy      2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504

Db       743 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 684

Qy      2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

Db       683 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 624

Qy      2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       623 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 564

Qy      2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       563 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 507

Qy      2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744

Db       506 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 449

Qy      2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
```

```
Db       448 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 413
```

```
RESULT 13
AX207897
LOCUS       AX207897                3147 bp    DNA     linear   PAT 31-AUG-2001
DEFINITION  Sequence 1 from Patent WO0157194.
ACCESSION   AX207897
VERSION     AX207897.1  GI:15422495
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madison,E.L., Ong,E.O. and Yeh,J.C.
  TITLE     Nucleic acid molecules encoding transmembrane serine proteases, the
            encoded proteins and methods based thereon
  JOURNAL   Patent: WO 0157194-A 1 09-AUG-2001;
            CORVAS INTERNATIONAL, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .3147
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="Nucleotide encoding MTSP1"
     CDS             23. .>2589
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC60377.1"
                     /db_xref="GI:15422496"
                     /translation="MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNV
                     KKVEKHGPGRWVVLAAVLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFV
                     DAYENSNSTEFVSLASKVKDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFS
                     IPQHLVEEAERVMAEERVVMLPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGL
                     HARGVELMRFTTPGFPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLV
                     TVYNTLSPMEPHALVQLCGTYPPSYNLTFHSSQNVLLITLITNTERRHPGFEATFFQL
                     PRMSSCGGRLRKAQGTFNSPYYPGHYPPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGV
                     PAGTCPKDYVEINGEKYCGERSQFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSS
                     DPCPGQFTCRTGRCIRKELRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVC
                     DSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVVT
                     CTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQARVVGGTDADEG
                     EWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFLGLHDQS
                     QRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFP
                     AGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFLSGGV
                     DSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTGV"
BASE COUNT      654 a    952 c    958 g    583 t
ORIGIN
```

```
  Query Match           60.6%;  Score 1883.2;  DB 6;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104

Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164

Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224

Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284

Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304
```

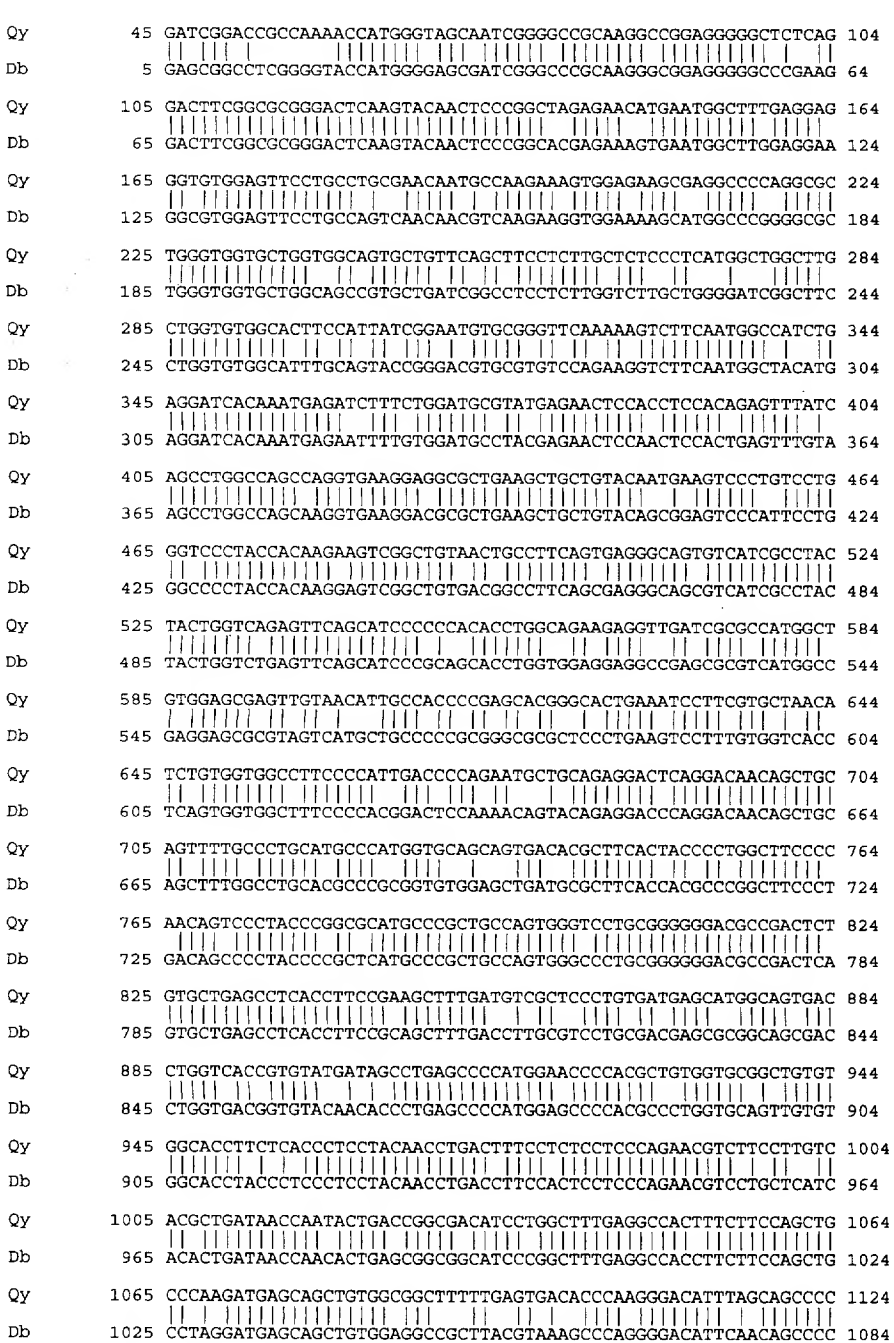
```
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
             ||||||||||||||||||||||||||| | ||||||||||||||||||| ||||||||
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
             |||||| ||  ||   | || ||  |     |||||  | |||   |||  | | |||
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
             || |  | || |   | |  |||  |||      |||    | ||  | |||  || |
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
             ||  | ||||||||||  |||   ||| |||  | |
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 14
AX207945
LOCUS       AX207945               3147 bp    DNA     linear   PAT 31-AUG-2001
DEFINITION  Sequence 49 from Patent WO0157194.
ACCESSION   AX207945
VERSION     AX207945.1  GI:15422543
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Madison,E.L., Ong,E.O. and Yeh,J.C.
  TITLE     Nucleic acid molecules encoding transmembrane serine proteases, the
            encoded proteins and methods based thereon
  JOURNAL   Patent: WO 0157194-A 49 09-AUG-2001;
            CORVAS INTERNATIONAL, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .3147
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             <1865. .2590
                     /note="protease domain of MTSP1"
                     /codon_start=1
                     /protein_id="CAC60383.1"
                     /db_xref="GI:15422544"
                     /translation="VVGGTDADEGEWPWQVSLHALGQGHICGASLISPNWLVSAAHCY
                     IDDRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLE
                     LEKPAEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQT
                     TCENLLPQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQ
                     RNKPGVYTRLPLFRDWIKENTGV"
BASE COUNT      654 a    952 c    958 g    583 t
ORIGIN

  Query Match             60.6%;  Score 1883.2;  DB 6;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
              || ||| |       |||||||| ||| |||||| |||||||| |||||||||| |  ||
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
              |||||||||||||||||||||||||||||||  |||||  ||||||||| |||||
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
              || |||||||||||||| |  ||||| | |||||| ||||| ||||  |||||  |||||
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
              ||||||||||||  || |||||| || || |||||||| |||  ||    |  |||||
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
              |||||||||| || || || ||| | ||||| || || || |||||||||||| |  ||
Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
              ||||||||||||||| ||| ||||||| || |||||||| |||||| |||||| |
Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
              |||||||||| |||||||||| |||||||||||||||||| | ||||||  |||||
Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
              || |||||||||| |||||||||| || |||||||| |||||||| ||||||||||||
Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
              ||||||| |||||||||||||| | |||||||  || ||||  || |||| ||||||
Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
              | |||||| || || |  |||| || || || || |  ||||| |||| ||| | ||
Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
              || |||||||| |||||||  ||| ||| ||   | |||||||| ||||||||||||||
Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
              || |||| |||||| |||| |||  |  ||| |||||||| || || |||||||||
Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCT 824
              ||||  |||||||| || |||||||||||||||||  |||||||||||||||||||||
Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
              |||||||||||||||||||  |||||||  | ||  |||| || ||||  |||| |||
Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
              |||||  || |||||  | |  |||||||||||||||  ||||||  |||||| |||||
Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
              |||||||  | |  |||||||||||||||||  |||| |||||||||||||||  ||  ||
Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
              || |||||||||| |||||  |||||  |||||||||||||||  ||||||||||||
Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
              || | |||||||||||||||  |||  ||  ||  | ||||  |||||||  | |||||||
Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
              |||||  |||||||||||||| ||||||||  |||||||||||||  ||  ||||||||||
Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
              |||| | | |||||||| |||||||| |||||||  ||| |||| |||  ||| || | |
Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
              ||| ||||| |||||||||  |||||||||  |||||||| |||||||| |||||||||
Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
              ||||| || || | |||||||||| |||||| ||||| | |||||| || ||||| ||
Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
```

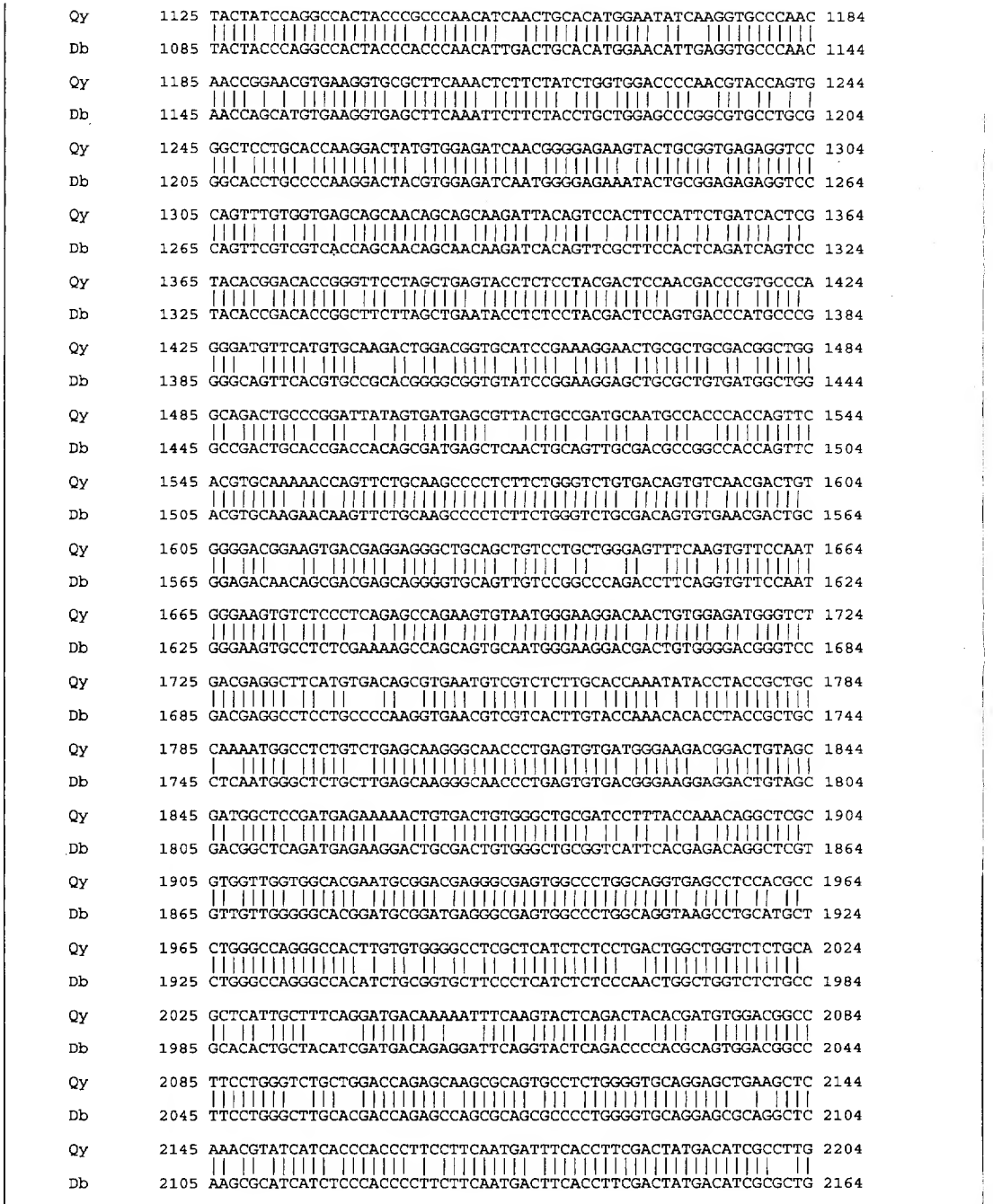
```
Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504

Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 15
AF057145

```
LOCUS       AF057145               3147 bp    mRNA    linear   PRI 22-SEP-2000
DEFINITION  Homo sapiens serine protease TADG15 mRNA, complete cds.
ACCESSION   AF057145
VERSION     AF057145.1  GI:10257389
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3147)
  AUTHORS   Tanimoto,H., Underwood,L.J., Wang,Y., Shigemasa,K., Parmley,T.H.
            and O'Brien,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-1998) Biochemistry, University of Arkansas for
            Medical Sciences, 4301 W. Markham St., Little Rock, AR 72205, USA
FEATURES             Location/Qualifiers
     source          1. .3147
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             23. .2590
                     /note="associated with human ovarian cancer"
                     /codon_start=1
                     /product="serine protease TADG15"
                     /protein_id="AAG15395.1"
                     /db_xref="GI:10257390"
                     /translation="MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNV
                     KKVEKHGPGRWVVLAAVLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFV
                     DAYENSNSTEFVSLASKVKDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFS
                     IPQHLVEEAERVMAEERVVMLPPRARSLKSFVVTSVVAFPTDSKTVQRTQDNSCSFGL
                     HARGVELMRFTTPGFPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLV
                     TVYNTLSPMEPHALVQLCGTYPPSYNLTFHSSQNVLLITLITNTERRHPGFEATFFQL
                     PRMSSCGGRLRKAQGTFNSPYYPGHYPPNIDCTWNIEVPNNQHVKVSFKFFYLLEPGV
                     PAGTCPKDYVEINGEKYCGERSQFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSS
                     DPCPGQFTCRTGRCIRKELRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVC
                     DSVNDCGDNSDEQGCSCPAQTFRCSNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVVT
                     CTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQARVVGGTDADEG
                     EWPWQVSLHALGQQHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFLGLHDQS
                     QRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFP
                     AGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFLSGGV
                     DSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENTGV"
BASE COUNT      654 a    952 c    958 g    583 t
ORIGIN
```

```
  Query Match           60.6%;  Score 1883.2;  DB 9;  Length 3147;
  Best Local Similarity  81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;
```





```
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324

Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384

Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444

Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504

Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


Search completed: February 21, 2004, 05:28:26
Job time : 11176 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:03:32 ; Search time 795 Seconds
(without alignments)
10546.484 Million cell updates/sec

Title: US-09-900-751-1
Perfect score: 3106
Sequence: 1 catggtagacggctgcccgg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3106	100.0	3106	24	AAD37039	Epithin gene. Uni
	2	1889.8	60.8	3149	21	AAA88493	Human matriptase c
	3	1884.8	60.7	3142	22	AAF28099	Human membrane-typ
	4	1883.2	60.6	3147	20	AAX87815	Tumour antigen der
	5	1883.2	60.6	3147	22	AAD13113	Human membrane-typ
	6	1883.2	60.6	3147	22	AAD13155	Human membrane-typ
	7	1883.2	60.6	3147	22	AAH23601	Human TADG-15 codi
c	8	1883.2	60.6	3147	22	AAH23609	Human TADG-15 anti
	9	1883.2	60.6	3147	24	AAL53444	Type II transmembr
	10	1883.2	60.6	3147	24	AAL53445	Type II transmembr
	11	1883.2	60.6	3147	25	ABZ58500	Transmembrane seri
	12	1883.2	60.6	3147	25	ABZ58501	Transmembrane seri
	13	1883.2	60.6	3147	25	ABZ22450	Human membrane-typ
	14	1883.2	60.6	3147	25	ABZ22451	Human MTSP1 protea
	15	1883.2	60.6	3147	25	AAD47180	Human membrane-typ
	16	1883.2	60.6	3147	25	AAD47181	Human membrane-typ
	17	1883.2	60.6	3147	25	AAD47225	Human membrane-typ
	18	1872.8	60.3	3152	22	AAH57431	Human intestine ce
	19	1872.8	60.3	3159	21	AAA37657	Human peptidase, H
c	20	1793.8	57.8	3115	22	ABA08672	Human membrane-typ
c	21	1793.8	57.8	3115	22	AAH99569	Human protein enco
	22	1762.2	56.7	2955	21	AAA88492	Human matriptase (
	23	1559.2	50.2	2756	21	AAC77957	Human cancer assoc
	24	1532.4	49.3	3112	24	ABS76526	cDNA encoding huma
	25	1530.8	49.3	2900	22	AAH23602	Human SNC-19 codin
	26	1530.8	49.3	2900	24	ABK84260	Human cDNA differe
	27	1498.8	48.3	3413	23	AAS85629	DNA encoding novel
	28	701.2	22.6	1553	24	ABK30240	Human G-protein-co
	29	556.4	17.9	726	24	AAD36928	Human matriptase o
	30	551.8	17.8	723	18	AAT79128	Human serine prote
	31	379.4	12.2	2086	23	AAS85628	DNA encoding novel
	32	367	11.8	1458	23	AAS85625	DNA encoding novel
	33	306.8	9.9	434	24	ABK39442	DNA encoding lung
	34	306.8	9.9	434	25	ACA11771	Human lung cancer
	35	306.8	9.9	434	25	ACA02957	Lung cancer therap
	36	302.6	9.7	429	24	ABL82332	Human ovarian canc
	37	279.2	9.0	387	23	AAS85626	DNA encoding novel
	38	265	8.5	591	23	AAS58827	cDNA #1503 encodin
	39	246.8	7.9	393	25	ABX40789	Bovine EST associa
	40	242.4	7.8	2801	24	ABQ61196	Human PRO618 encod
	41	240.8	7.8	2672	22	AAD13117	Human membrane-typ
	42	240.8	7.8	3104	22	AAD13116	Human membrane-typ
	43	239.2	7.7	1327	20	AAZ34034	Human EST DNA35597
	44	239.2	7.7	1327	21	AAC78495	Human EST DNA35597
	45	239.2	7.7	1327	21	AAC58237	Human EST (express

ALIGNMENTS

RESULT 1
AAD37039
ID AAD37039 standard; DNA; 3106 BP.
XX
AC AAD37039;
XX
DT 21-AUG-2002 (first entry)
XX
DE Epithin gene.
XX
KW Transgenic; transgenic animal; pharmacological therapy; gene therapy;
KW phenotype modulation; genetic disease; epithin; gene; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 63..2630
FT /*tag= a
FT /product= "Epithin protein"
FT misc_feature 2332..2465
FT /*tag= b
FT /note= "Sequence flanking Neo insert in
FT targetting construct"
FT misc_feature 2466..2465
FT /*tag= c
FT /note= "Sequence deleted in targetting construct"
FT misc_feature 2466..3033
FT /*tag= d
FT /note= "Sequence flanking Neo insert in
FT targetting construct"
XX

PN WO200203787-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21427.
XX
PR 06-JUL-2000; 2000US-216109P.
PR 06-JUL-2000; 2000US-216251P.
PR 06-JUL-2000; 2000US-216258P.
PR 06-JUL-2000; 2000US-216768P.
PR 10-JUL-2000; 2000US-217449P.
PR 10-JUL-2000; 2000US-217450P.
PR 10-JUL-2000; 2000US-217660P.
PR 27-JUL-2000; 2000US-221491P.
PR 27-JUL-2000; 2000US-221669P.
PR 27-JUL-2000; 2000US-221670P.
PR 07-AUG-2000; 2000US-223170P.
PR 07-AUG-2000; 2000US-223172P.
PR 07-AUG-2000; 2000US-223460P.
PR 26-OCT-2000; 2000US-244037P.
PR 26-OCT-2000; 2000US-244111P.
PR 26-JUN-2001; 2001US-301217P.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Allen KD, Leviten MW;
XX
DR WPI; 2002-154853/20.
DR P-PSDB; AAE23083.
XX
PT Novel non-human transgenic animal, preferably transgenic mice
PT comprising disruption in target gene, e.g., tryptase gene, useful for
PT identifying an agent that modulates expression or function of target
PT gene -
XX
PS Example 2; Fig 2A; 74pp; English.
XX
CC The present invention relates to non-human transgenic animals preferably
CC transgenic mice comprising disruption in target gene such as tryptase
CC gene. The invention also relates to compositions and methods relating
CC to the characterisation of gene functions. The transgenic animals are
CC useful for identifying an agent that modulates the expression or function
CC of a target. They are useful for identifying an agent that modulates a
CC phenotype associated with a disruption in tryptase genes or limulus
CC clotting factor protease-like genes by administering an agent to the
CC transgenic animal and determining whether the agent modulates the
CC phenotype where the agent has effect on decreased body weight, decreased
CC thymus weight, decreased thymus to body weight ratio, increased pre-pulse
CC inhibition, significant decrease in their response latency to the hot
CC plate test or a decreased response threshold to metrazol. Agents that
CC modulate the expression, function or activity of the target gene are
CC useful for treating a disorder associated with a mutation in tryptase
CC gene or in limulus clotting factor protease-like gene. The transgenic
CC animals are useful for testing the efficacy of proposed genetic and
CC pharmacological therapies for human genetic diseases. They are useful
CC as models for diseases, disorders or conditions associated with
CC phenotypes relating to a disruption in a target and to identify drugs,
CC pharmaceuticals, therapies and interventions which may be effective in
CC treating a disease or other phenotypic characteristics of the animal.
CC The present sequence is epithin gene. This sequence is used in the
CC exemplification of the invention.
XX
SQ Sequence 3106 BP; 692 A; 862 C; 897 G; 655 T; 0 other;

Query Match 100.0%; Score 3106; DB 24; Length 3106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240

Qy        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300

Qy        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360

Qy        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420

Qy        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480

Qy        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540

Qy        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600

Qy        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660

Qy        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720

Qy        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780

Qy        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840

Qy        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900

Qy        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960

Qy        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020

Qy       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080

Qy       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140

Qy       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

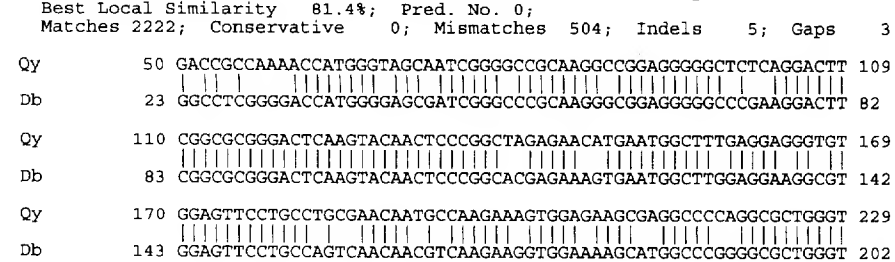
```
Db       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
Qy       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
Db       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
Qy       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
Db       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
Qy       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
Db       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
Qy       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
Db       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
Qy       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
Db       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
Qy       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
Db       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
Qy       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
Db       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
Qy       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
Db       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
Qy       1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
Db       1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
Qy       1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
Db       1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
Qy       1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
Db       1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
Qy       1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
Db       1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
Qy       1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
Db       1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
Qy       1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
Db       1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
Qy       2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100
Db       2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100
Qy       2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160
Db       2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160
Qy       2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220
Db       2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220
Qy       2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280
Db       2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280
```

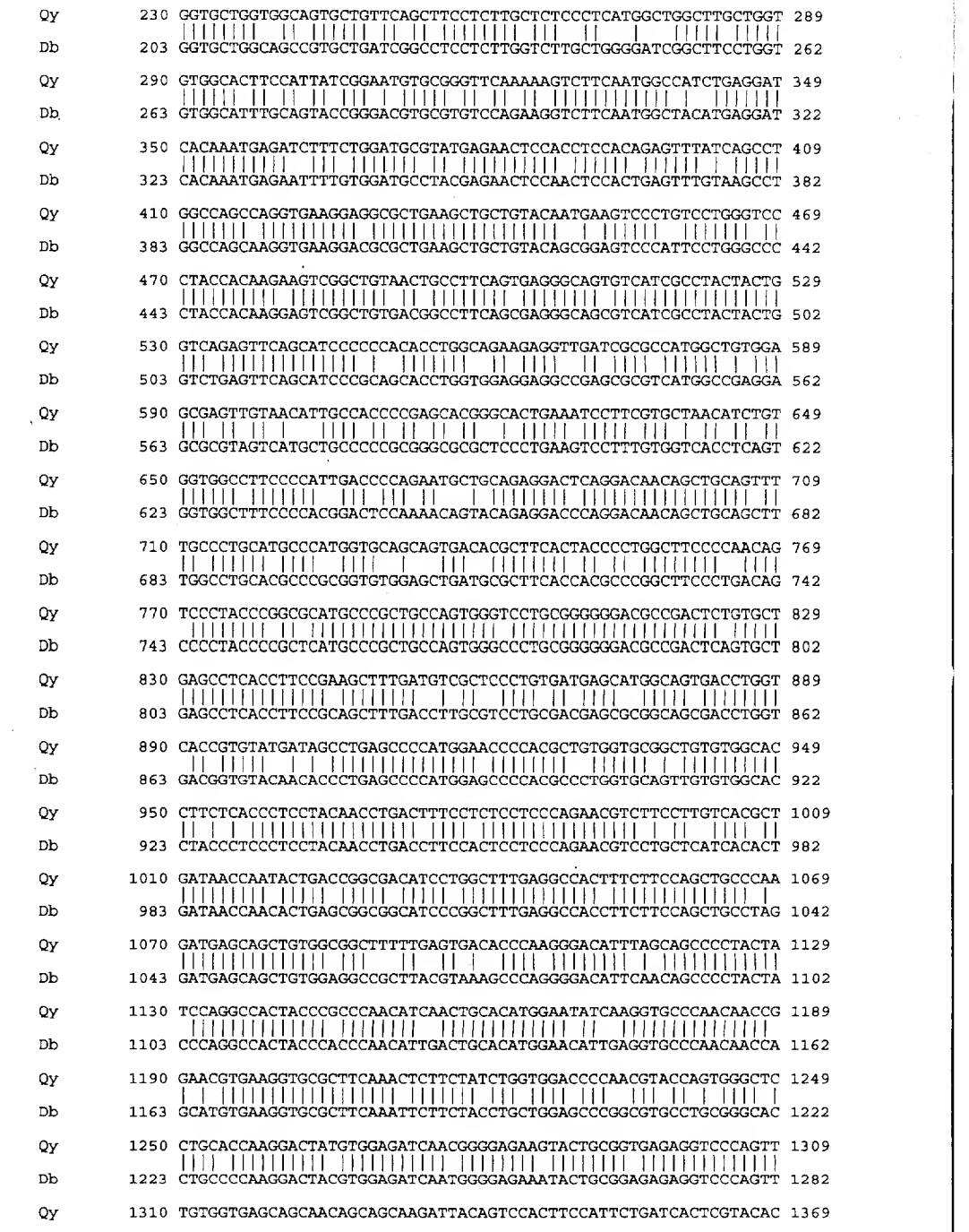
```
Qy       2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
Db       2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
Qy       2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400
Db       2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400
Qy       2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
Db       2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
Qy       2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
Db       2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
Qy       2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
Db       2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
Qy       2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
Db       2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
Qy       2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
Db       2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
Qy       2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
Db       2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
Qy       2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820
Db       2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820
Qy       2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880
Db       2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880
Qy       2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940
Db       2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940
Qy       2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000
Db       2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000
Qy       3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060
Db       3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060
Qy       3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
Db       3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
```

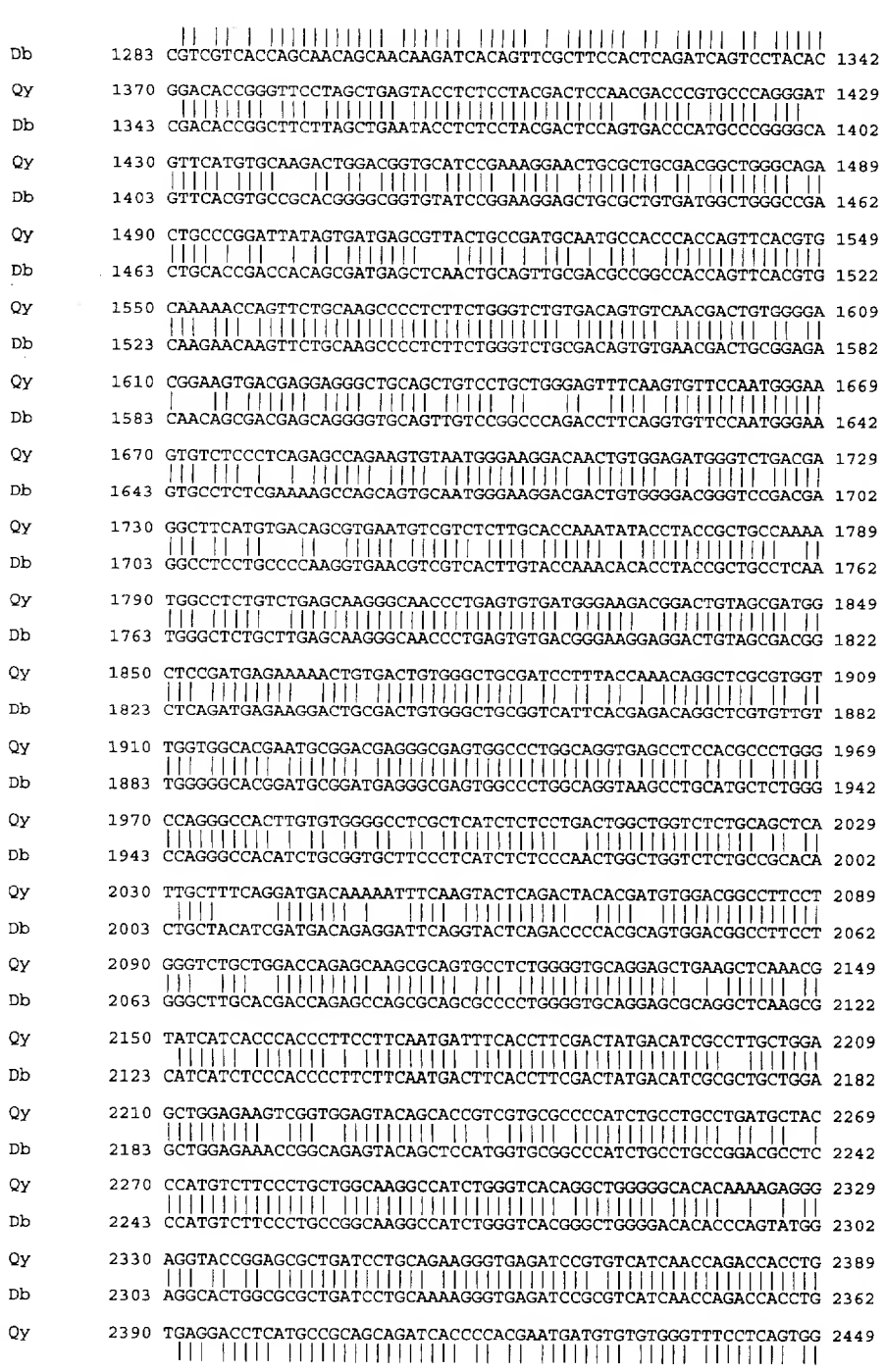
```
RESULT 2
AAA88493
ID   AAA88493 standard; cDNA; 3149 BP.
XX
AC   AAA88493;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Human matriptase cDNA.
XX
KW   Matriptase; serine protease; human; breast cancer; pre-malignancy;
KW   actinic keratosis; leukoplakia; Barrett's epithelium;
KW   columnar metaplasia; ulcerative colitis; bowenoid papulosis;
KW   adenomatous colorectal polyp; Qyerat erythroplasia;
KW   vulvar intraepithelial neoplasia; tumour; metastasis; therapy; ss.
```

XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 36..2603
FT /*tag= c
XX
PN WO200053232-A1.
XX
PD 14-SEP-2000.
XX
PF 10-MAR-2000; 2000WO-US06111.
XX
PR 12-MAR-1999; 99US-0124006.
XX
PA (GEOU) UNIV GEORGETOWN.
XX
PI Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;
XX
DR WPI; 2000-594268/56.
DR P-PSDB; AAB19552.
XX
PT Treating malignancies, premalignant and pathologic conditions in a
PT subject, comprises administering matriptase modulating agent -
XX
PS Claim 9; Fig 15; 116pp; English.
XX
CC The present sequence is that of cDNA coding for the full-length
CC form (see AAB19521) of human matriptase, a trypsin-like protease.
CC cDNA encoding a truncated form of matriptase is given in AAA88492.
CC Either form can be incorporated into a vector and used in a claimed
CC method for making recombinant matriptase. The zymogen (inactive)
CC form of matriptase is a single-chain protein. The active 2-chain
CC form strongly interacts with fragments of a Kunitz-type serine
CC protease inhibitor (hepatocyte growth factor activator inhibitor,
CC HAI-1) to form SDS-stable complexes. In breast cancer cells,
CC matriptase is present mainly as the uncomplexed form. Only the
CC complexed matriptase in detected in human milk. The invention is
CC directed to a method of detecting a malignancy or a pre-malignant
CC lesion in breast or other tissue by detecting the presence of
CC single- or 2-chain forms of matriptase in the tissue. The
CC object is to inhibit tumour onset, tumour growth and metastasis.
CC Malignancies and pre-malignant conditions characterised by
CC expression of the zymogen or activated form of matriptase are
CC treated by administering an inhibitor of matriptase, especially a
CC Bowman-Birk inhibitor. The pre-malignant condition is atypical
CC ductal hyperplasia of the breast, actinic keratosis, leukoplakia,
CC Barrett's epithelium of the esophagus, ulcerative colitis,
CC adenomatous colorectal polyps, erythroplasia of the Queyrat,
CC Bowen's disease, bowenoid papulosis, vulvar intraepithelial
CC neoplasia or dysplatic changes to the cervix. The invention also
CC provides methods for in vivo or in vitro diagnosis of malignancy
CC or pre-malignant lesion, and methods of identifying matriptase
CC modulators, including activators and inhibitors.
XX
SQ Sequence 3149 BP; 647 A; 959 C; 962 G; 581 T; 0 other;

Query Match 60.8%; Score 1889.8; DB 21; Length 3149;
Best Local Similarity 81.4%; Pred. No. 0;
Matches 2222; Conservative 0; Mismatches 504; Indels 5; Gaps 3;



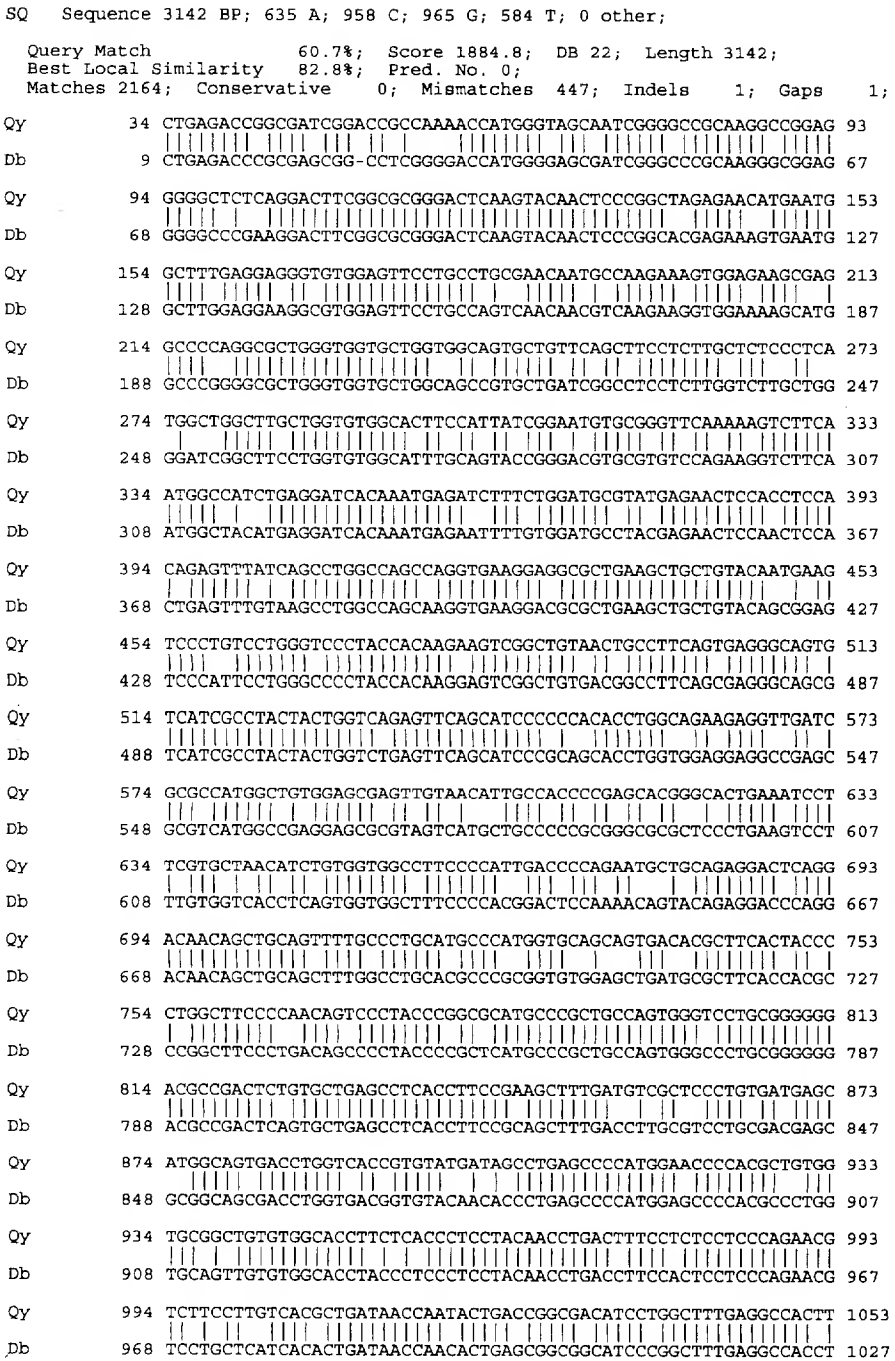


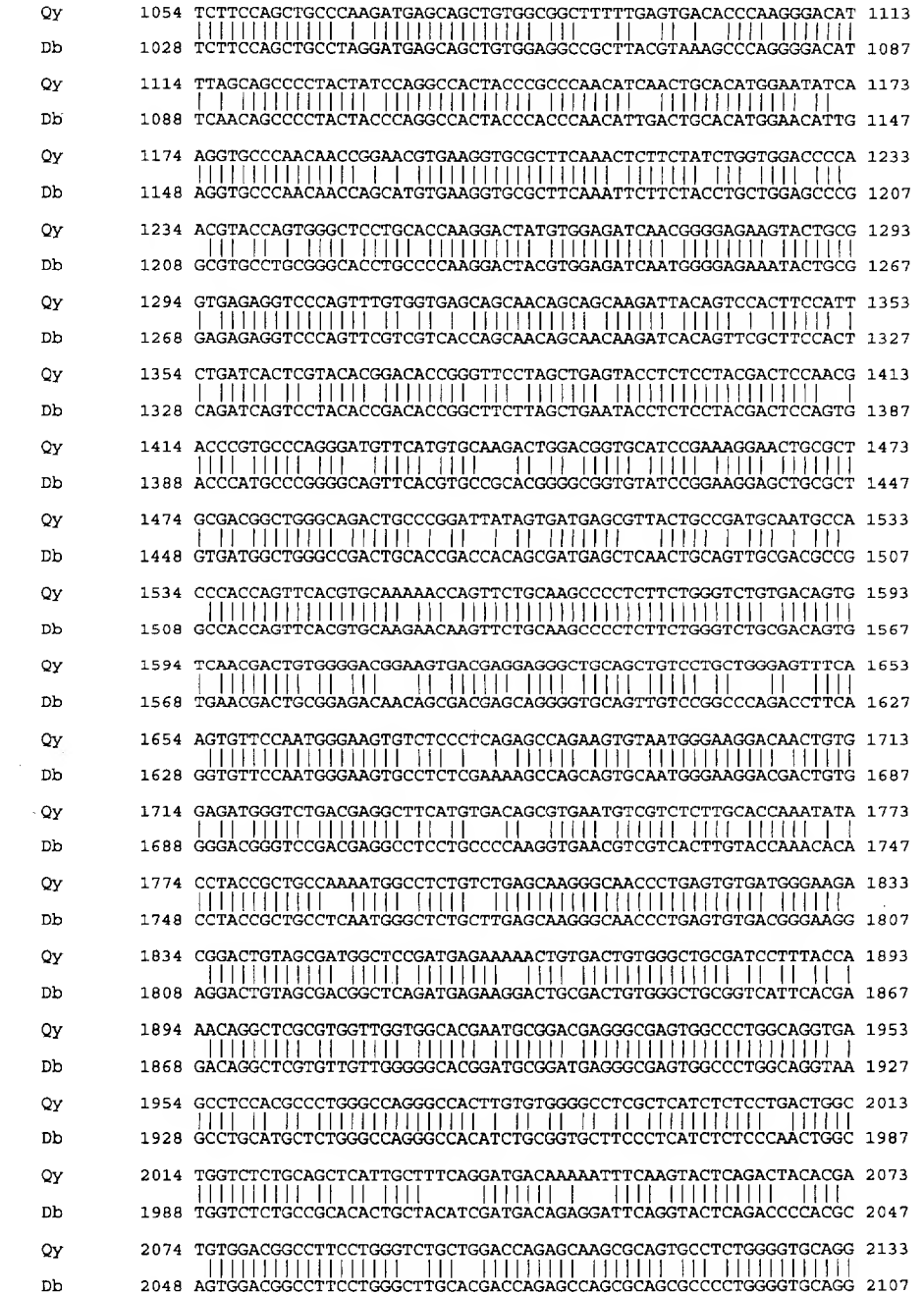


RESULT 3
AAF28099
ID AAF28099 standard; cDNA; 3142 BP.
XX
AC AAF28099;
XX
DT 06-JUN-2001 (first entry)
XX
DE Human membrane-type serine protease MT-SP1 coding sequence.
XX
KW Human; membrane-type serine protease; MT-SP1; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 37..2602
FT /*tag= a
FT /product= "MT-SP1"
XX
PN WO200123524-A2.
XX
PD 05-APR-2001.
XX
PF 02-OCT-2000; 2000WO-US27250.
XX
PR 30-SEP-1999; 99US-0410362.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Craik CS, Takeuchi T, Shuman M;
XX
DR WPI; 2001-245002/25.
DR P-PSDB; AAB35465.
XX
PT New nucleic acid encoding a membrane type serine protease, useful for
PT the diagnosis, prognosis and treatment of cancer, particularly
PT metastatic cancers -
XX
PS Claim 4; Fig 1; 102pp; English.
XX
CC The present invention provides the protein and coding sequences for the
CC novel human membrane-type serine protease MT-SP1. Increased expression of
CC this protein is associated with cancer, and so the sequences can be used
CC in cancer diagnosis and the identification of treatments. The present
CC sequence is the MT-SP1 coding sequence.
XX

SQ Sequence 3142 BP; 635 A; 958 C; 965 G; 584 T; 0 other;

Query Match 60.7%; Score 1884.8; DB 22; Length 3142;
Best Local Similarity 82.8%; Pred. No. 0;
Matches 2164; Conservative 0; Mismatches 447; Indels 1; Gaps 1;



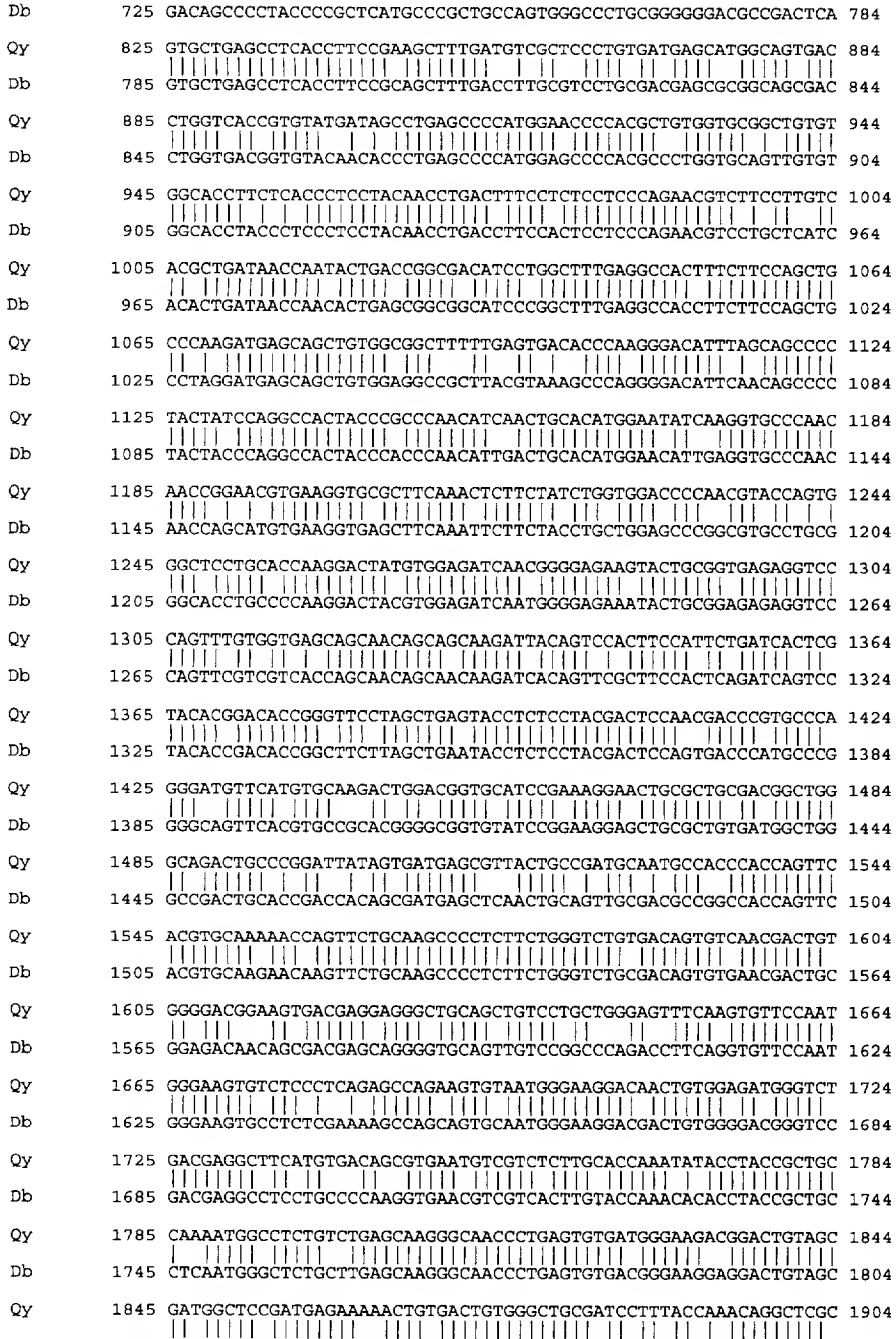


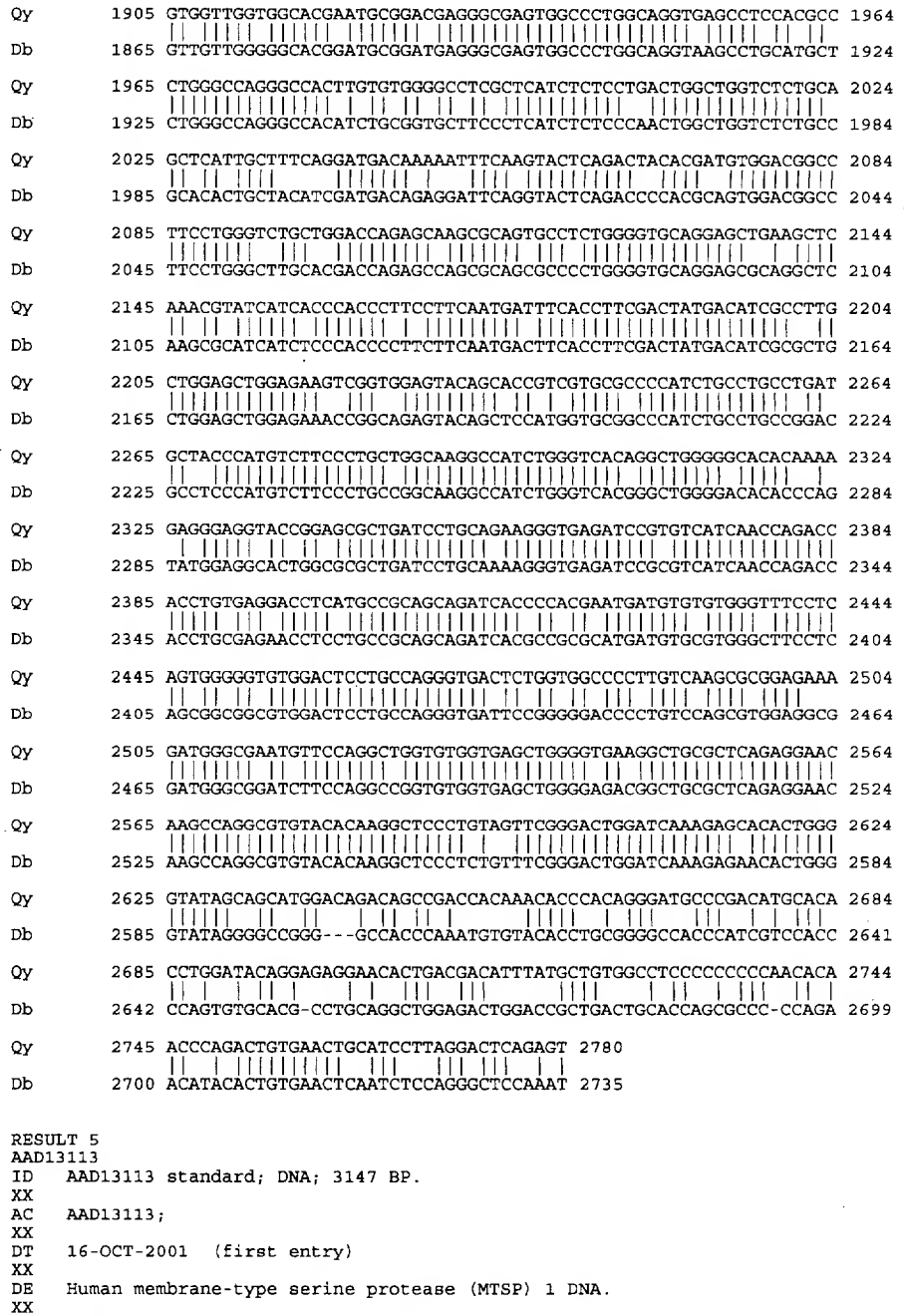
```
Qy       2134 AGCTGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATG 2193
Db       2108 AGCGCAGGCTCAAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATG 2167

Qy       2194 ACATCGCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCT 2253
Db       2168 ACATCGCGCTGCTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCT 2227

Qy       2254 GCCTGCCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGG 2313
Db       2228 GCCTGCCGGACGCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGG 2287

Qy       2314 GGCACACAAAAGAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCA 2373
Db       2288 GACACACCCAGTATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCA 2347

Qy       2374 TCAACCAGACCACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTG 2433
Db       2348 TCAACCAGACCACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCG 2407

Qy       2434 TGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAA 2493
Db       2408 TGGGCTTCCTCAGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCA 2467

Qy       2494 GCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCG 2553
Db       2468 GCGTGGAGGCGGATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCG 2527

Qy       2554 CTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAG 2613
Db       2528 CTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAG 2587

Qy       2614 AGCACACTGGGGTATAGCAGCATGGACAGACA 2645
Db       2588 AGAACACTGGGGTATAGGGGCCGGGGCACCCA 2619
```

RESULT 4
AAX87815
ID AAX87815 standard; cDNA; 3147 BP.
XX
AC AAX87815;
XX
DT 09-NOV-1999 (first entry)
XX
DE Tumour antigen derived gene-15 (TADG-15) cDNA.
XX
KW Tumour antigen derived gene-15; TADG-15; serine protease; human;
KW breast cancer; ovary cancer; carcinoma; diagnosis; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   23..2590
FT   /*tag= a
XX
PN WO9942120-A1.
XX
PD 26-AUG-1999.
XX
PF 18-FEB-1999; 99WO-US03436.
XX
PR 20-FEB-1998; 98US-0027337.
XX
PA (UYAR-) UNIV ARKANSAS.
XX
PI O'Brien TJ, Tanimoto H;
XX
DR WPI; 1999-527418/44.
DR P-PSDB; AAY06671.
XX
PT A new extracellular serine protease for diagnosis of neoplastic
PT disease
XX
PS Claim 2; Fig 9; 71pp; English.
XX
CC This is the nucleotide sequence of an isolated cDNA that codes for
CC an extracellular serine protease, termed tumour antigen derived
CC gene-15 protein (see AAY06671), that is overexpressed in breast and
CC ovarian carcinomas. The TADG-15 gene can be used as a diagnostic
CC and therapeutic target in ovarian carcinoma and other carcinomas
CC including breast, prostate, lung and colon. The TADG-15 cDNA was
CC isolated from ovarian carcinoma by PCR using primers directed to
CC conserved areas of the serine protease family. The invention
CC also provides: a vector that is capable of expressing DNA encoding
CC TADG-15 protein; host cells selected from bacterial cells
CC (especially Escherichia coli), mammalian cells, plant cells and
CC insect cells; and a method of detecting expression of TADG-15
CC protein using a hybridisation probe.
XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 20; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

```
Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
```





RESULT 5
AAD13113
ID AAD13113 standard; DNA; 3147 BP.
XX
AC AAD13113;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human membrane-type serine protease (MTSP) 1 DNA.
XX
KW Human; transmembrane serine protease; membrane-type serine protease;

```
KW   MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
KW   lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KW   matriptase; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             23..2590
FT                   /*tag=  a
FT                   /product= "Human transmembrane serine protease (MTSP) 1"
XX
PN   WO200157194-A2.
XX
PD   09-AUG-2001.
XX
PF   02-FEB-2001; 2001WO-US03471.
XX
PR   03-FEB-2000; 2000US-0179982.
PR   18-FEB-2000; 2000US-0183542.
PR   22-JUN-2000; 2000US-0213124.
PR   26-JUL-2000; 2000US-0220970.
PR   08-SEP-2000; 2000US-0657986.
PR   22-SEP-2000; 2000US-0234840.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Madison EL,  Ong EO,  Yeh J;
XX
DR   WPI; 2001-488877/53.
DR   P-PSDB; AAE06930.
XX
PT   Novel single chain polypeptide comprising protease domain of type-II
PT   membrane-type serine protease or its catalytically active portion
PT   useful for treating and preventing cancer and tumor  -
XX
PS   Claim 14; Page 191-195; 256pp; English.
XX
CC   The invention relates to transmembrane serine proteases and their
CC   corresponding nucleotides and the protease domain of a type-II
CC   membrane-type serine protease (MTSP). MTSP is useful for identifying
CC   compounds that modulate or inhibits its proteolytic activity and for
CC   formulating a medicament for treating neoplastic disease. MTSP and
CC   its corresponding nucleotides are useful in preventing or treating
CC   tumours or cancers such as lung carcinoma, colon adenocarcinoma and
CC   ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
CC   is useful as a diagnostic marker for tumour development, growth and/or
CC   progression and as immunogens to generate antibodies that specifically
CC   bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
CC   gene therapy. The present sequence is a DNA encoding human MTSP1
CC   protein (also called matriptase).
XX
SQ   Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

  Query Match           60.6%;  Score 1883.2;  DB 22;  Length 3147;
  Best Local Similarity  81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
              ||  |||  |       |||||||| ||| ||||||  |||||||| ||||||||| |  ||
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG  64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
              |||||||||||||||||||||||||||||||||  |||||  |||||||||| |||||
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
              ||  |||||||||||||  |  |||||  | ||||||  |||||  ||||  |||||  |||||
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
              ||||||||||||  ||  ||||||  ||  ||  ||||||||  |||  ||    |  |||||
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
              ||||||||||  ||  ||  |||  |  |||||  ||  ||  ||  ||||||||||  |  ||
Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
              ||||||||||||||||  |||  |||||||  ||  |||||||||  ||||||  ||||||  |
Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
              ||||||||||||  ||||||||||  ||||||||||||||||||||  |  ||||||  |||||
Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
              ||  ||||||||||||  ||||||||||  ||  ||||||||  ||||||||  ||||||||||
Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
              ||||||||  ||||||||||||||  |  |||||||  ||  ||||  ||  ||||  ||||||
Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
              |  ||||||  ||  ||  |    ||||  ||  ||  ||  ||  |  |||||  |||||  |||  |  ||
Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
              ||  ||||||||  |||||||  |||  |||  ||    |  ||||||||  ||||||||||||||
Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
              ||  ||||  ||||||  ||||  ||||  |  |||  ||||||||  ||  ||  ||||||||
Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCT 824
              ||||  ||||||||  ||  ||||||||||||||||||  ||||||||||||||||||||
Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
              ||||||||||||||||||||  ||||||||  |  ||  ||||  ||  ||||  |||||  |||
Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
              |||||  ||  |||||  |  |  ||||||||||||||  |||||||  ||||||  |  |||||
Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
              |||||||  |  ||||||||||||||||  ||||  ||||||||||||||  |  ||  ||
Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
              ||  ||||||||||  |||||  |||||  ||||  ||||||||||||  |||||||||||
Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
              ||  |  ||||||||||||||||  |||    ||  ||  |  ||||  ||||||||  |  |||||||
Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
              |||||  ||||||||||||||  |||||||  ||||||||||||  ||  ||||||||||
Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
              ||||  |  |||||||||  |||||||  ||||||  |||  ||||  |||  ||  |  |
Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
              |||  |||||  ||||||||||  ||||||||||  ||||||||  ||||||||  |||||||||
Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
              |||||  ||  ||  |  ||||||||||  ||||||  |||||  |  ||||||  ||  |||||  ||
Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
```

```
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484

Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544

Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604

Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664

Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724

Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784

Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844

Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904

Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964

Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024

Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084

Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144

Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204

Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264

Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324

Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384

Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444

Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504

Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 6
AAD13155

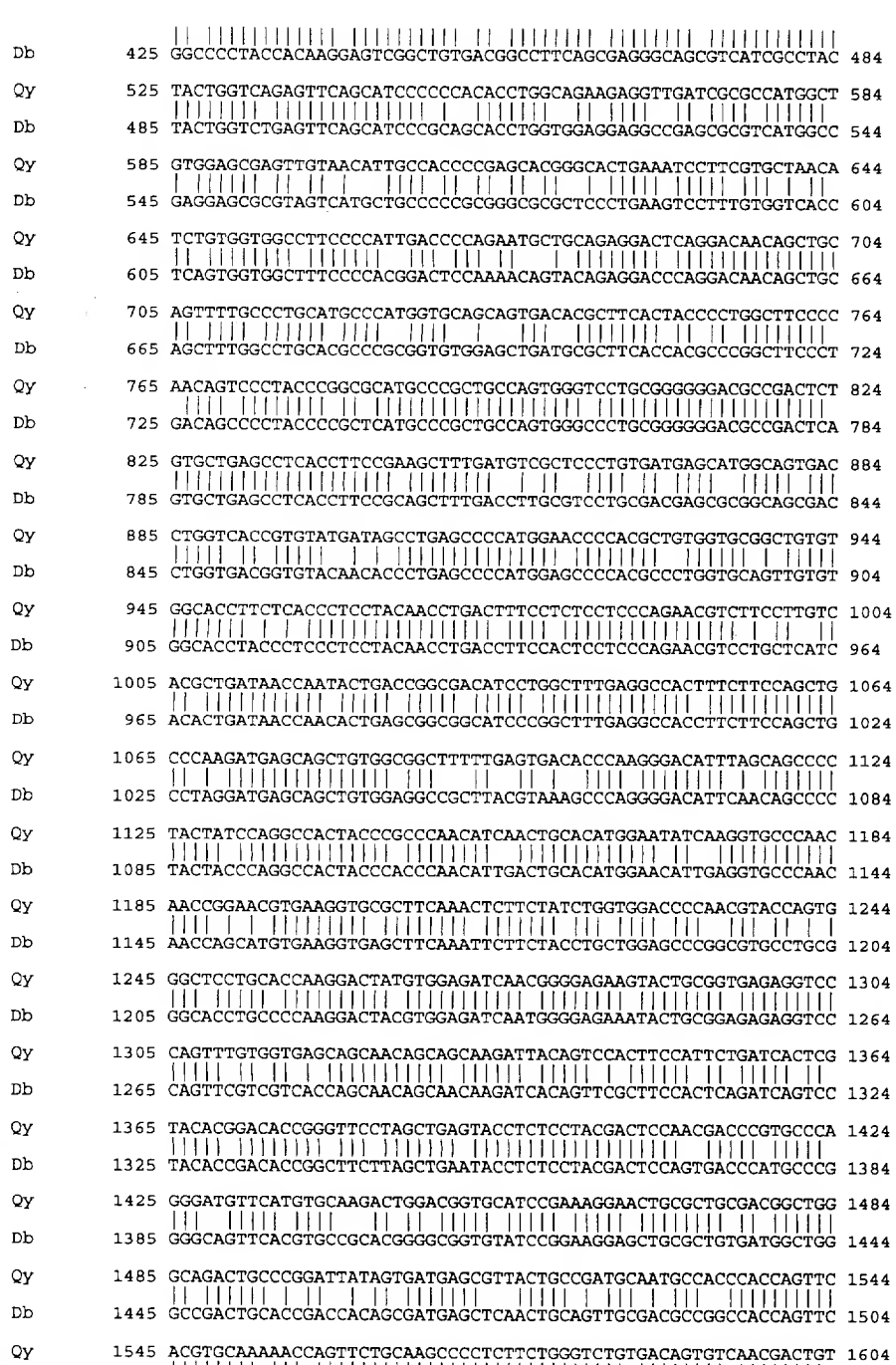
```
ID   AAD13155 standard; DNA; 3147 BP.
XX
AC   AAD13155;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human membrane-type serine protease (MTSP) 1 protease domain DNA.
XX
KW   Human; transmembrane serine protease; membrane-type serine protease;
KW   MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
KW   lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KW   matriptase; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1865..2590
FT                   /*tag=  a
FT                   /product= "Human transmembrane serine protease
FT                   (MTSP) 1 protease domain"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200157194-A2.
XX
PD   09-AUG-2001.
XX
PF   02-FEB-2001; 2001WO-US03471.
XX
PR   03-FEB-2000; 2000US-0179982.
PR   18-FEB-2000; 2000US-0183542.
PR   22-JUN-2000; 2000US-0213124.
PR   26-JUL-2000; 2000US-0220970.
PR   08-SEP-2000; 2000US-0657986.
PR   22-SEP-2000; 2000US-0234840.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Madison EL,  Ong EO,  Yeh J;
XX
DR   WPI; 2001-488877/53.
DR   P-PSDB; AAE06936.
XX
PT   Novel single chain polypeptide comprising protease domain of type-II
PT   membrane-type serine protease or its catalytically active portion
PT   useful for treating and preventing cancer and tumor  -
XX
PS   Example 6; Page 225-227; 256pp; English.
```

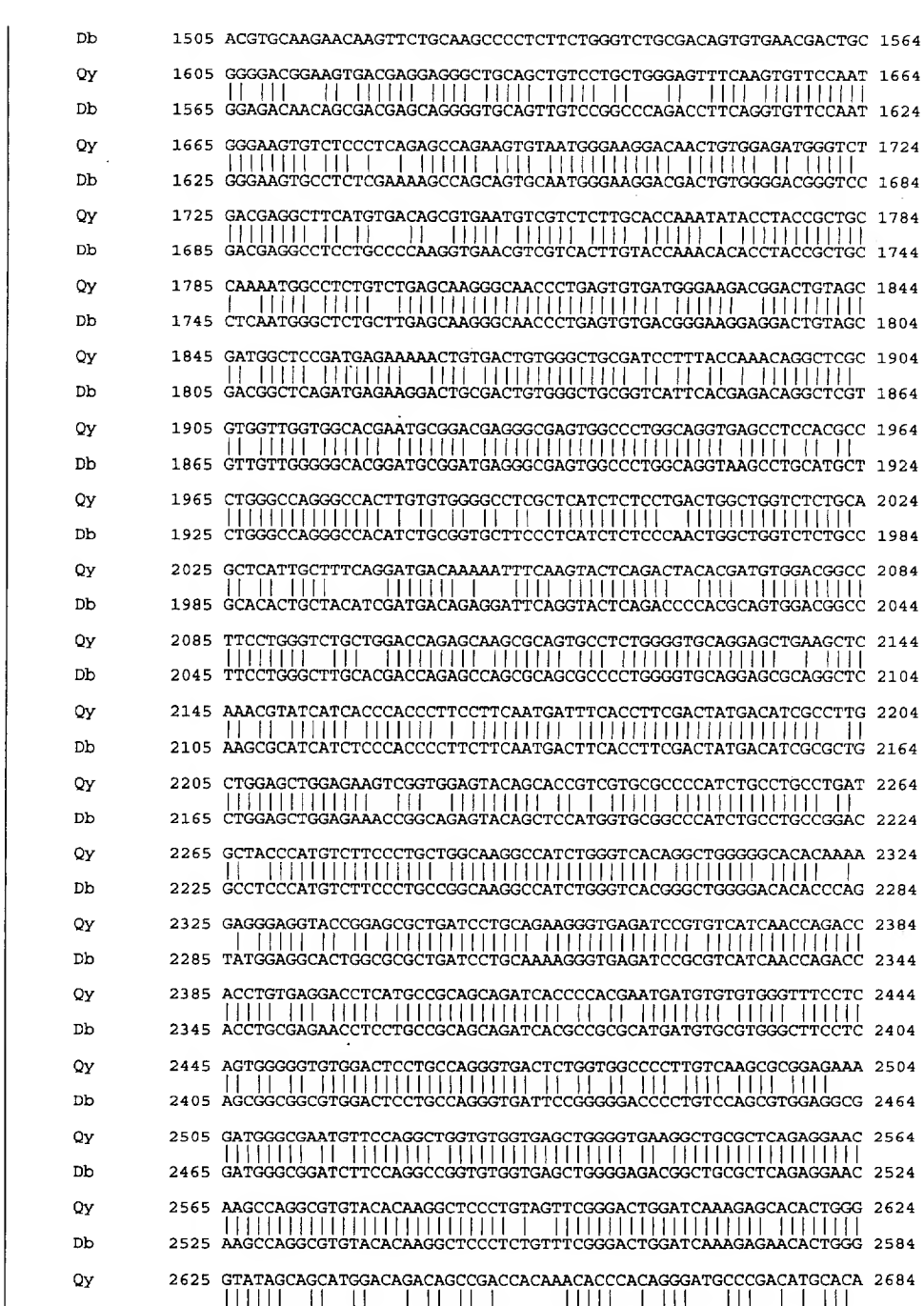
XX
CC The invention relates to transmembrane serine proteases and their
CC corresponding nucleotides and the protease domain of a type-II
CC membrane-type serine protease (MTSP). MTSP is useful for identifying
CC compounds that modulate or inhibits its proteolytic activity and for
CC formulating a medicament for treating neoplastic disease. MTSP and
CC its corresponding nucleotides are useful in preventing or treating
CC tumours or cancers such as lung carcinoma, colon adenocarcinoma and
CC ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
CC is useful as a diagnostic marker for tumour development, growth and/or
CC progression and as immunogens to generate antibodies that specifically
CC bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
CC gene therapy. The present sequence is a DNA encoding protease domain of
CC human MTSP1 protein (also called matriptase).
XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 22; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

```
Qy        45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
             || ||| |       |||||||| ||| |||||| |||||||| ||||||||| |   ||
Db         5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy       105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
             ||||||||||||||||||||||||||||||||  |||||  |||||||||| |||||
Db        65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy       165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
             || |||||||||||| |  |||||  | ||||||  |||||  ||||  |||||  |||||
Db       125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy       225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
             ||||||||||||  || ||||||  || || ||||||||  |||  ||    |  |||||
Db       185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy       285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
             ||||||||||| || || || ||| | ||||| || || || ||||||||||| |  ||
Db       245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy       345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
             |||||||||||||||  ||| |||||| || |||||||||  |||||  ||||||  |
Db       305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy       405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
             |||||||||||  |||||||||  ||||||||||||||||||  | ||||||  |||||
Db       365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy       465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
             || |||||||||||  ||||||||||  || ||||||||  ||||||||  ||||||||||||
Db       425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy       525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
             ||||||||  ||||||||||||||  |||||||  ||  ||||  ||  |||| ||||||
Db       485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy       585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
             | ||||||  ||  || |   ||||  ||  ||  ||  | |||||  |||||  ||| | ||
Db       545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy       645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
             ||  |||||||  |||||||  |||  |||  ||    | ||||||||  ||||||||||||||
Db       605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy       705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
             ||  ||||  ||||||  |||  ||||  |  |||  ||||||||  ||  ||||||||||
Db       665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy       765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
               ||||  ||||||||  ||  ||||||||||||||||||  ||||||||||||||||||||
Db       725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784

Qy       825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
             |||||||||||||||||||  ||||||||  | ||  |||| ||  ||||  ||||| |||
Db       785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy       885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
             |||||  ||  |||||  |  | ||||||||||||||  ||||||||  ||||||  | |||||
Db       845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy       945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
             |||||||  |  ||||||||||||||||||  ||||  ||||||||||||||||  |  ||  ||
Db       905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy      1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
             ||  ||||||||||  |||||  |||||  |||||||||||||||  ||||||||||
Db       965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy      1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
             ||  | ||||||||||||||||  |||  ||  ||  |  ||||  ||||||||  | |||||||
Db      1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy      1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
             |||||  ||||||||||||||  ||||||||  ||||||||||||  ||  ||||||||||||
Db      1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy      1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
             |||| | | |||||||||  |||||||  ||| |||| |||  ||| || | |
Db      1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy      1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
             |||  |||||  ||||||||||  ||||||||||  ||||||||  ||||||||  |||||||||
Db      1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy      1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
             |||||  ||  ||  |  ||||||||||  ||||||  |||||  |  ||||||  ||  |||||  ||
Db      1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy      1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
             |||||  |||||||||  |||  |||||||  ||||||||||||||||||||  |||||  |||||
Db      1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy      1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
             |||  |||||  ||||    ||  ||  |||||  |||||  |||||  ||||||||  ||  ||||||
Db      1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy      1485 GCAGACTGCCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
             ||  ||||||  |  ||  |  ||  |||||||  |||||  |  |||  |  |||  ||||||||||
Db      1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy      1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
             ||||||||  |||  ||||||||||||||||||||||||||||||||  ||||||||  ||||||||
Db      1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy      1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
             ||  |||    ||  ||||||  ||||  |||||  |||||  ||    ||  ||||  ||||||||||
Db      1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy      1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
             ||||||||  |||  |  ||||||  ||||  ||||||||||||  ||||||  ||  |||||
Db      1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy      1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
             ||||||||  ||  ||    ||  |||||  ||||  ||||||  |  ||||||||||||
Db      1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy      1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
             |  |||||  |||||  ||||||||||||||||||||||||||  ||||||  ||||||||||
Db      1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy      1845 GATGGCTCCGATGAGAAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
             ||  |||||  ||||||||  ||||  ||||||||||||||  ||  ||  ||  |||||||||
Db      1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy      1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
```

```
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024

Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084

Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144

Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204

Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264

Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324

Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384

Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444

Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504

Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 7
AAH23601
ID   AAH23601 standard; DNA; 3147 BP.
XX
AC   AAH23601;
XX
DT   03-AUG-2001  (first entry)
XX
DE   Human TADG-15 coding sequence.
XX
KW   Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;
KW   tumour antigen-derived gene 15; extracellular serine protease; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             23..2590
FT                   /*tag=  a
FT                   /product= "Human TADG-15"
XX
PN   WO200129056-A1.
XX
PD   26-APR-2001.
XX
PF   20-OCT-2000; 2000WO-US29095.
XX
PR   20-OCT-1999;   99US-0421213.
XX
PA   (UYAR-) UNIV ARKANSAS.
XX
PI   O'Brien TJ,  Tanimoto H;
XX
DR   WPI; 2001-381031/40.
DR   P-PSDB; AAB98500.
XX
PT   Novel extracellular serine protease, termed tumor antigen-derived gene
PT   15 protein overexpressed in carcinomas and DNA encoding it, for
PT   diagnosis, treatment, prevention of cancer, particularly breast,
PT   ovarian cancer -
XX
PS   Claim 2; Fig 2; 130pp; English.
XX
CC   The present sequence is the coding sequence for human tumour
CC   antigen-derived gene 15 (TADG-15) protein. TADG-15 is an extracellular
CC   serine protease. It was found that TADG-15 is over-expressed in ovarian
CC   tumours. TADG-15 protein or its fragments of 9-20 residues that lack
CC   TADG-15 protease activity are useful for vaccinating an individual
CC   against TADG-15, having, suspected of having or at risk of getting
CC   cancer. Furthermore, the TADG-15 gene can be used as a diagnostic or
CC   therapeutic target in cancer.
XX
SQ   Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

  Query Match           60.6%;  Score 1883.2;  DB 22;  Length 3147;
  Best Local Similarity  81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104

Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164

Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224

Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284

Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404

Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464

Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
```





```
Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 8
AAH23609/c
ID AAH23609 standard; RNA; 3147 BP.
XX
AC AAH23609;
XX
DT 03-AUG-2001 (first entry)
XX
DE Human TADG-15 antisense RNA sequence.
XX
KW TADG-15; cytostatic; vaccine; ovarian tumour; cancer; human; antisense;
KW tumour antigen-derived gene 15; serine protease; ss.
XX
OS Homo sapiens.
XX
PN WO200129056-A1.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000WO-US29095.
XX
PR 20-OCT-1999; 99US-0421213.
XX
PA (UYAR-) UNIV ARKANSAS.
XX
PI O'Brien TJ, Tanimoto H;
XX
DR WPI; 2001-381031/40.
XX
PT Novel extracellular serine protease, termed tumor antigen-derived gene
PT 15 protein overexpressed in carcinomas and DNA encoding it, for
PT diagnosis, treatment, prevention of cancer, particularly breast,
PT ovarian cancer -
XX
PS Example 8; Page 102-103; 130pp; English.
XX
CC The present invention relates to human tumour antigen-derived gene 15
CC (TADG-15) protein and coding sequence (see AAH23601 and AAB98500).
CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is
CC over-expressed in ovarian tumours. TADG-15 protein or its fragments of
CC 9-20 residues that lack TADG-15 protease activity are useful for
CC vaccinating an individual against TADG-15, having, suspected of having or
CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a
CC diagnostic or therapeutic target in cancer. TADG-15 was cloned and
CC expressed in the opposite orientation such that an antisense RNA molecule
CC (the present sequence) was produced. The present sequence is useful for
CC hybridising to the complementary RNA in a cell and thereby inhibiting
CC translation of TADG-15 into protein.
XX
SQ Sequence 3147 BP; 583 A; 958 C; 952 G; 654 U; 0 other;

Query Match 60.6%; Score 1883.2; DB 22; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

```
Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db       3143 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 3084

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db       3083 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 3024

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db       3023 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 2964

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db       2963 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 2904

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db       2903 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 2844

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db       2843 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 2784

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db       2783 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 2724

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db       2723 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 2664

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db       2663 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 2604

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db       2603 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 2544

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db       2543 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 2484

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db       2483 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 2424

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db       2423 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 2364

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db       2363 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 2304

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db       2303 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 2244

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db       2243 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 2184

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db       2183 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 2124

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db       2123 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 2064

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db       2063 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 2004

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db       2003 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1944

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
```

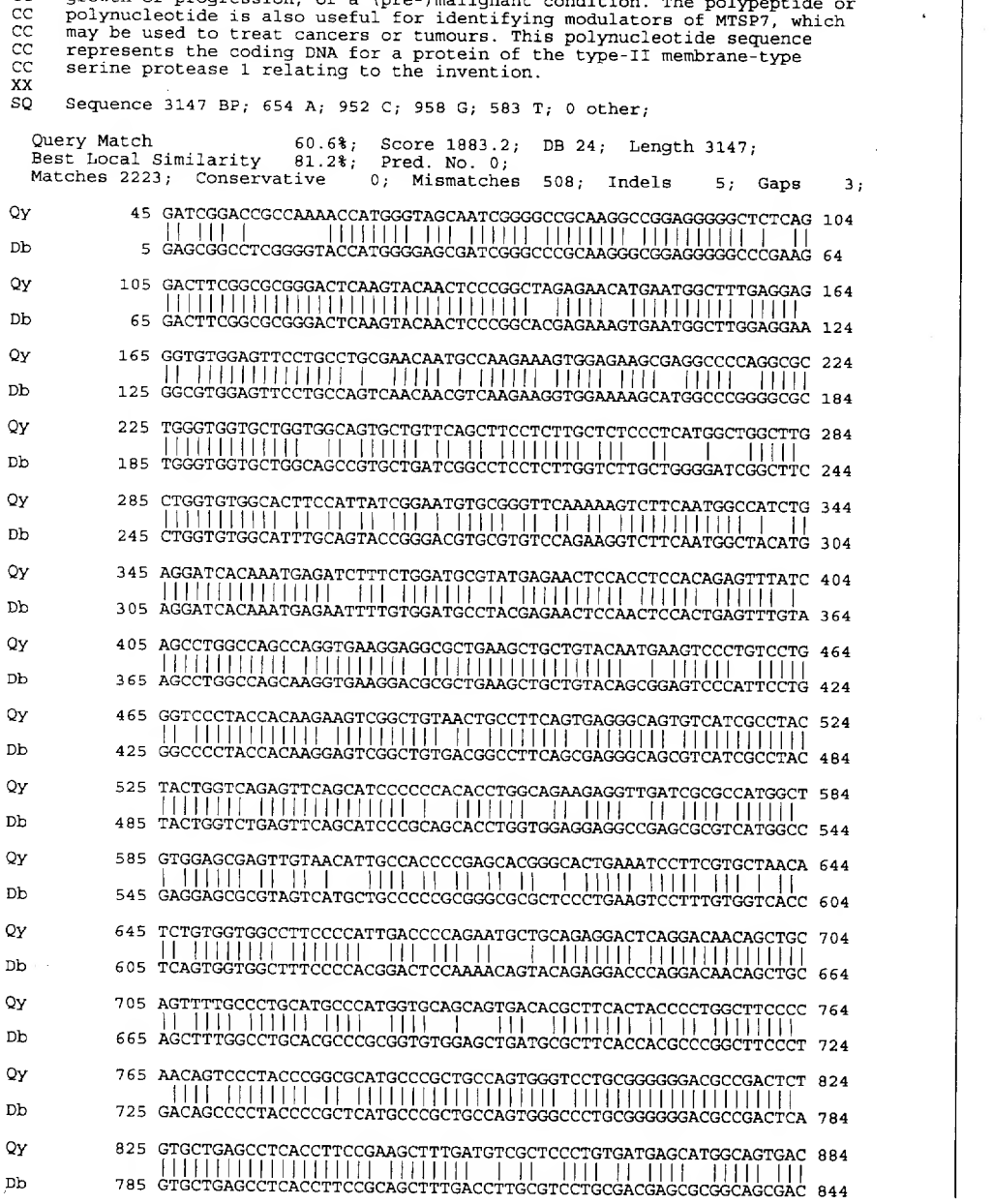
```
Db       1943 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1884
Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db       1883 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1824
Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db       1823 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1764
Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db       1763 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1704
Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db       1703 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1644
Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db       1643 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1584
Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db       1583 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1524
Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db       1523 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1464
Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db       1463 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1404
Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db       1403 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1344
Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db       1343 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1284
Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db       1283 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1224
Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db       1223 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1164
Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db       1163 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 1104
Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db       1103 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 1044
Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db       1043 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 984
Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db        983 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 924
Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db        923 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 864
Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
```

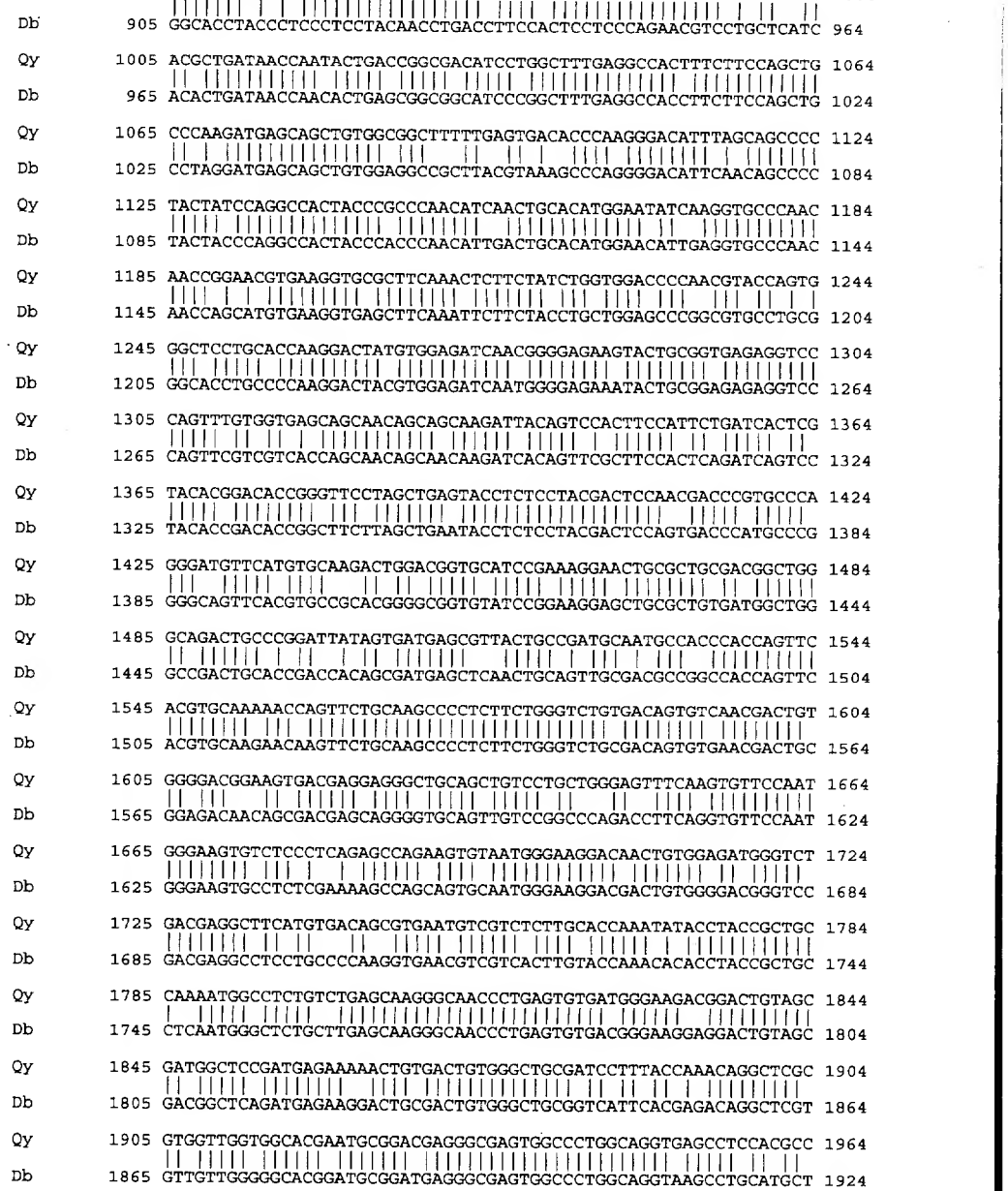
```
Db        863 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 804
Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db        803 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 744
Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db        743 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 684
Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db        683 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 624
Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db        623 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 564
Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db        563 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 507
Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db        506 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 449
Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db        448 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 413
```

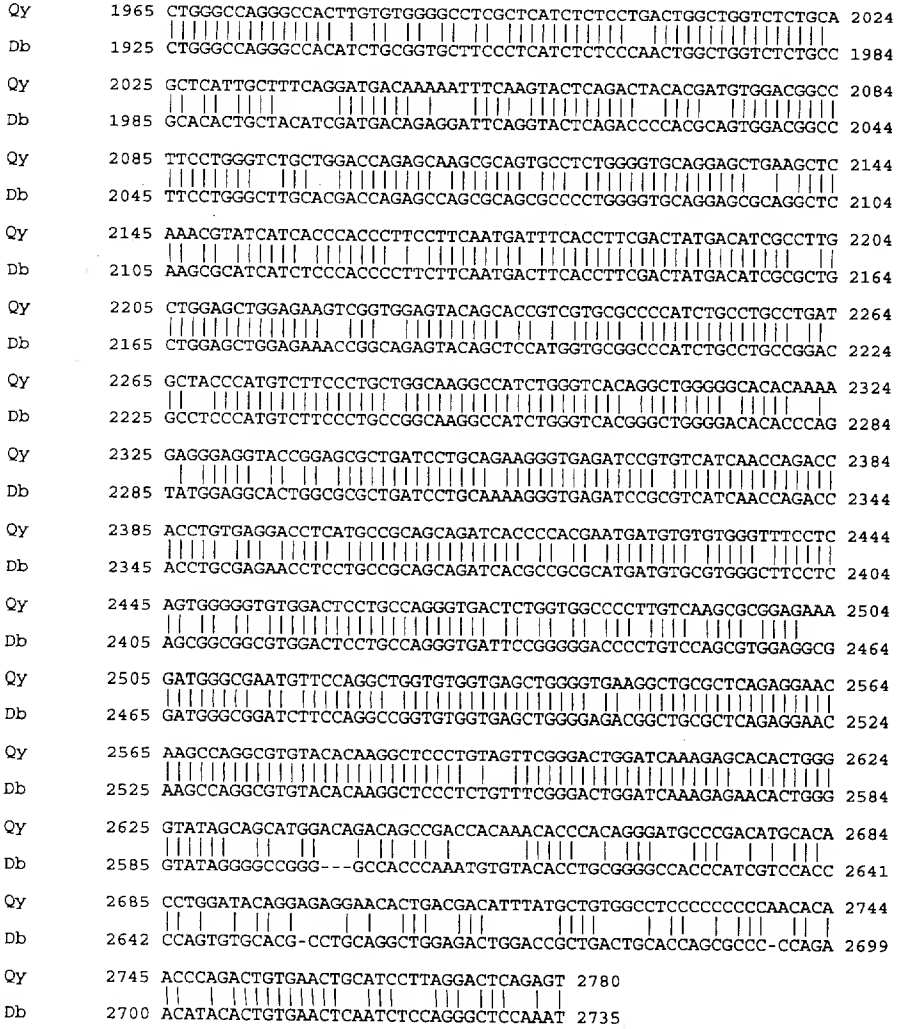
RESULT 9
AAL53444
ID   AAL53444 standard; DNA; 3147 BP.
XX
AC   AAL53444;
XX
DT   12-DEC-2002 (first entry)
XX
DE   Type II transmembrane serine protease 1 coding DNA SEQ ID No 1.
XX
KW   Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy;
KW   neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;
KW   malignant; enzyme; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200272786-A2.
XX
PD   19-SEP-2002.
XX
PF   13-MAR-2002; 2002WO-US07903.
XX
PR   13-MAR-2001; 2001US-275592P.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Madison EL, Ong EO;
XX
DR   WPI; 2002-732827/79.
DR   P-PSDB; AAO22929.
XX
PT   New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing
PT   neoplastic diseases, monitoring tumor progress or therapeutic
PT   effectiveness, or identifying MTSP7 modulators for treating tumors or
PT   cancers -
XX
PS   Disclosure; Page 168-172; 184pp; English.
XX
CC   The invention relates to a purified single or two-chain polypeptide,
CC   which comprises the protease domain of a type-II membrane-type serine
CC   protease 7 (MTSP7) or its catalytically active portion. The polypeptide
CC   comprising MTSP7 is useful for detecting or diagnosing a neoplastic

CC disease, a pre-malignant lesion, a malignancy or other pathologic
CC condition in a subject. This polypeptide is also useful for monitoring
CC tumour (e.g. tumuor of the breast, cervix, prostate, lung, ovary or
CC colon) progress and/or therapeutic effectiveness. The inhibitor of the
CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful
CC for treating or preventing a neoplastic disease, or tumour initiation,
CC growth or progression, or a (pre-)malignant condition. The polypeptide or
CC polynucleotide is also useful for identifying modulators of MTSP7, which
CC may be used to treat cancers or tumours. This polynucleotide sequence
CC represents the coding DNA for a protein of the type-II membrane-type
CC serine protease 1 relating to the invention.
XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 24; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;



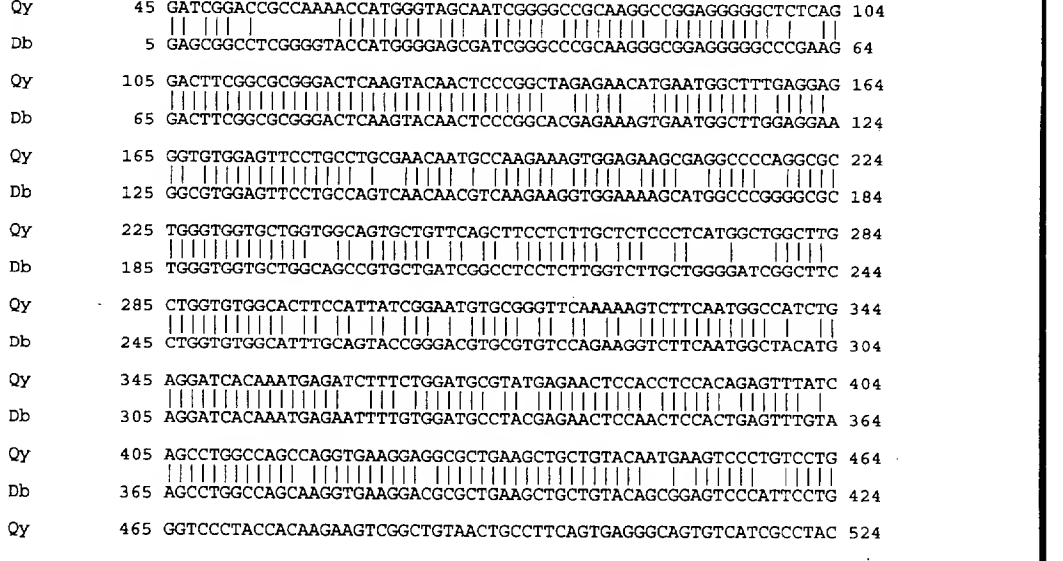


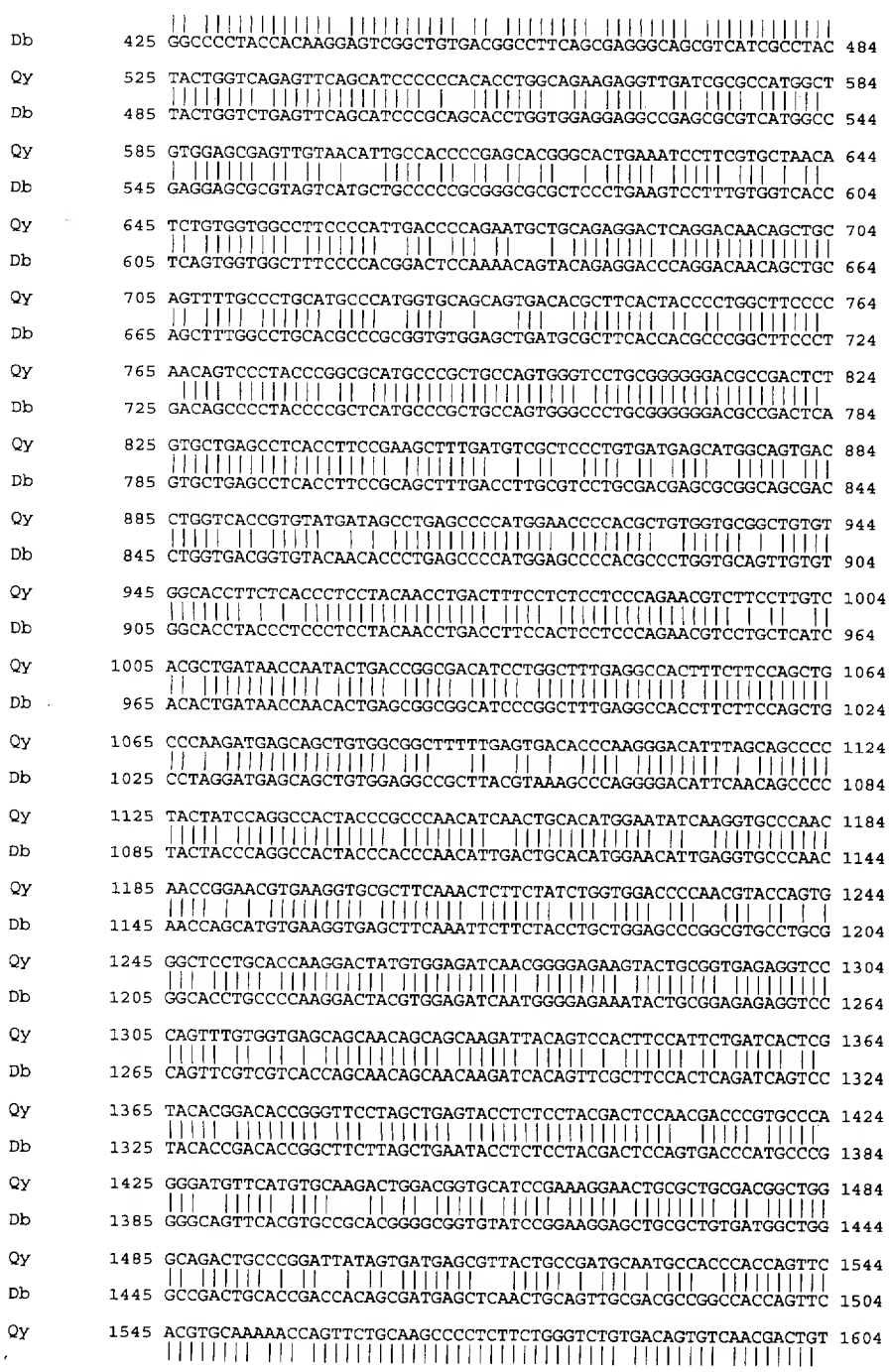


RESULT 10
AAL53445
ID AAL53445 standard; DNA; 3147 BP.
XX
AC AAL53445;
XX
DT 12-DEC-2002 (first entry)
XX
DE Type II transmembrane serine protease 1 domain DNA SEQ ID No 3.
XX
KW Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy;
KW neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;
KW malignant; enzyme; gene; ds.
XX
OS Homo sapiens.
XX

PN WO200272786-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US07903.
XX
PR 13-MAR-2001; 2001US-275592P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Ong EO;
XX
DR WPI; 2002-732827/79.
DR P-PSDB; AAO22930.
XX
PT New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing
PT neoplastic diseases, monitoring tumor progress or therapeutic
PT effectiveness, or identifying MTSP7 modulators for treating tumors or
PT cancers -
XX
PS Disclosure; Page 174-176; 184pp; English.
XX
CC The invention relates to a purified single or two-chain polypeptide,
CC which comprises the protease domain of a type-II membrane-type serine
CC protease 7 (MTSP7) or its catalytically active portion. The polypeptide
CC comprising MTSP7 is useful for detecting or diagnosing a neoplastic
CC disease, a pre-malignant lesion, a malignancy or other pathologic
CC condition in a subject. This polypeptide is also useful for monitoring
CC tumour (e.g. tumuor of the breast, cervix, prostate, lung, ovary or
CC colon) progress and/or therapeutic effectiveness. The inhibitor of the
CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful
CC for treating or preventing a neoplastic disease, or tumour initiation,
CC growth or progression, or a (pre-)malignant condition. The polypeptide or
CC polynucleotide is also useful for identifying modulators of MTSP7, which
CC may be used to treat cancers or tumours. This polynucleotide sequence
CC represents the coding DNA for a protein domain of the type-II membrane-
CC type serine protease 1 relating to the invention.
XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 24; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;




Db 425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484
Qy 525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db 485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544
Qy 585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db 545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604
Qy 645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664
Qy 705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db 665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724
Qy 765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db 725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784
Qy 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844
Qy 885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db 845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904
Qy 945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db 905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964
Qy 1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db 965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024
Qy 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084
Qy 1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144
Qy 1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db 1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204
Qy 1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAAATACTGCGGAGAGAGGTCC 1264
Qy 1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db 1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
Qy 1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db 1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
Qy 1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
Qy 1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db 1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504
Qy 1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564
Qy 1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db 1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624
Qy 1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744
Qy 1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db 1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804
Qy 1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864
Qy 1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db 1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924
Qy 1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984
Qy 2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044
Qy 2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104
Qy 2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164
Qy 2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224
Qy 2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284
Qy 2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db 2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344
Qy 2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db 2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404
Qy 2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db 2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464
Qy 2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524
Qy 2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db 2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584
Qy 2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db 2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

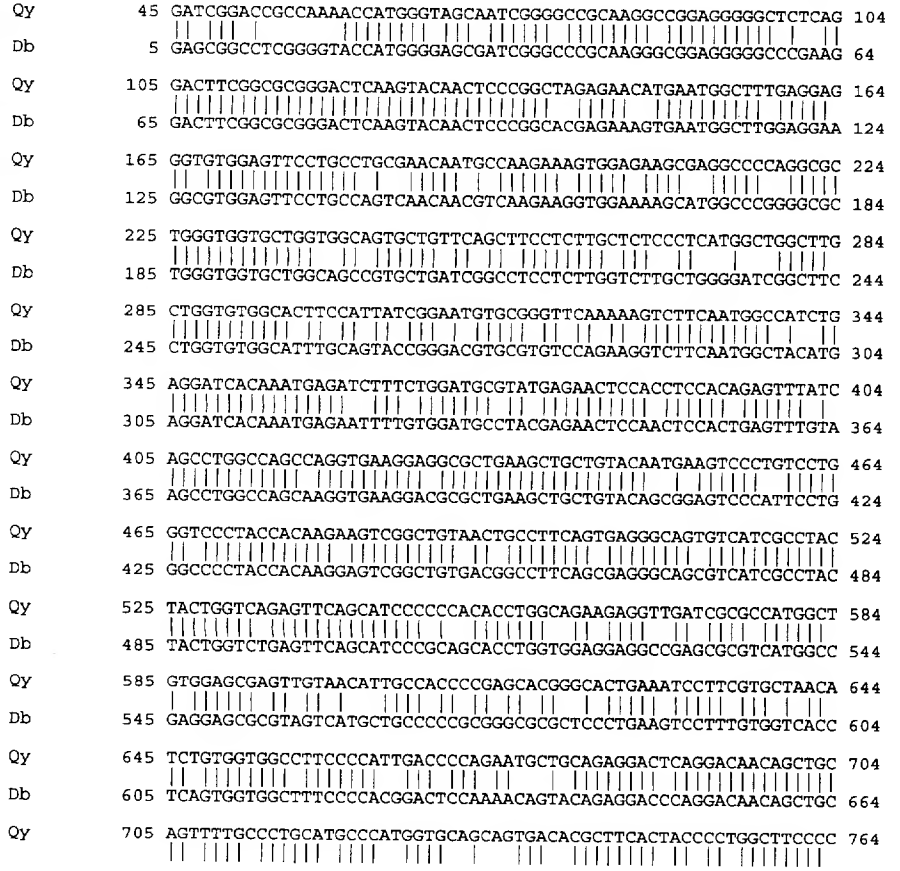
```
Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 11
ABZ58500
ID   ABZ58500 standard; cDNA; 3147 BP.
XX
AC   ABZ58500;
XX
DT   13-MAY-2003  (first entry)
XX
DE   Transmembrane serine protease 1 (MTSP1) cDNA.
XX
KW   Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;
KW   cytostatic; dermatological; cardiant; vulnerary; ophthalmological;
KW   gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             23..2590
FT                   /*tag=  a
FT                   /product= "MTSP1"
XX
PN   WO2003004681-A2.
XX
PD   16-JAN-2003.
XX
PF   03-JUL-2002; 2002WO-US21208.
XX
PR   03-JUL-2001; 2001US-302939P.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Madison EL,  Ong EO;
XX
DR   WPI; 2003-239207/23.
DR   P-PSDB; ABP72376.
XX
PT   New type-II membrane-type serine protease 20 polypeptides, useful for
PT   preparing a medicament for diagnosing, treating or preventing cancer,
PT   dermatological disorders, aberrant wound repairs or crest syndromes -
XX
PS   Disclosure; Page 194-198; 216pp; English.
XX
CC   The present sequence is that of cDNA encoding human type II
CC   transmembrane serine protease 1 (MTSP1).  The invention relates
CC   to novel human type II transmembrane serine protein 20 (MTSP20)
CC   (see ABP72374) and nucleic acids encoding it (see ABZ58499).
CC   Also claimed are methods of inhibiting tumour initiation, growth
CC   or progression by inhibiting MTSP20 activity, and of treating or
CC   preventing a disease or disorder associated with undesired and/or
CC   uncontrolled angiogenesis or neovascularisation, especially
CC   undesired angiogenesis associated with solid neoplasms, vascular
CC   malformations and cardiovascular disorders, chronic inflammatory
CC   diseases, aberrant wound repairs, circulatory disorders, crest
CC   syndromes, dermatological disorders and ocular disorders using
CC   an inhibitor of MTSP20.
XX
SQ   Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

  Query Match           60.6%;  Score 1883.2;  DB 25;  Length 3147;
  Best Local Similarity  81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104

Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164

Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224

Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284

Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404

Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464

Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524

Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584

Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644

Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704

Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764

Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCT 824

Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884

Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944

Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004

Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064

Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124

Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
```

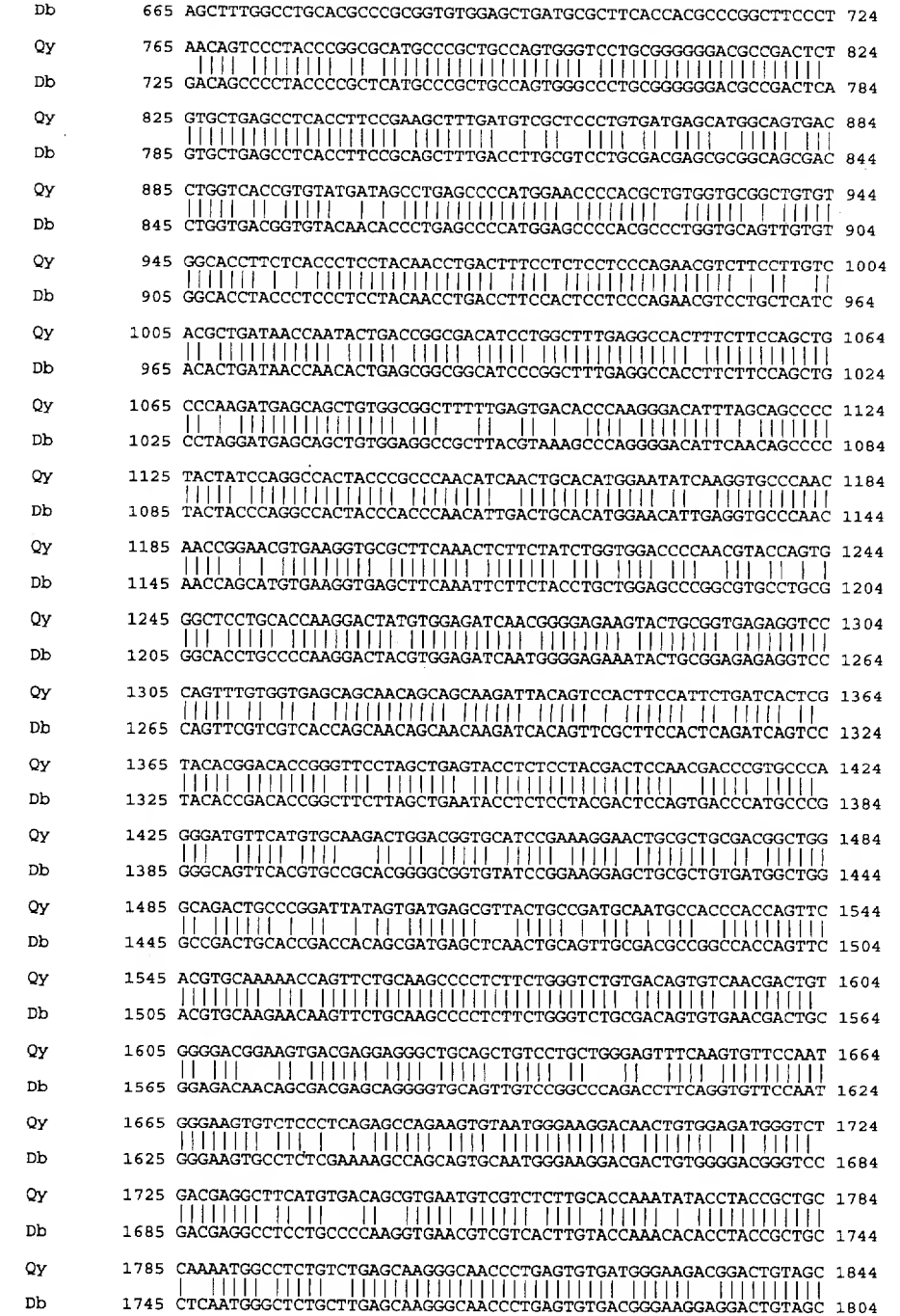
```
Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144
Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204
Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264
Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504
Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564
Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624
Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684
Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744
Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804
Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864
Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924
Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984
Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044
Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104
Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164
Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284
Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344
Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404
Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464
Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524
Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584
Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641
Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699
Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 12
ABZ58501
ID ABZ58501 standard; cDNA; 3147 BP.
XX
AC ABZ58501;
XX
DT 13-MAY-2003 (first entry)
XX
DE Transmembrane serine protease 1 (MTSP1) protein domain cDNA.
XX
KW Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;
KW cytostatic; dermatological; cardiant; vulnerary; ophthalmological;
KW gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1865..2590
FT   /*tag= a
FT   /product= "MTSP1 protease domain"
XX
PN WO2003004681-A2.
XX
PD 16-JAN-2003.
XX
PF 03-JUL-2002; 2002WO-US21208.
XX
PR 03-JUL-2001; 2001US-302939P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Ong EO;
XX
DR WPI; 2003-239207/23.
DR P-PSDB; ABP72377.

XX
PT New type-II membrane-type serine protease 20 polypeptides, useful for
PT preparing a medicament for diagnosing, treating or preventing cancer,
PT dermatological disorders, aberrant wound repairs or crest syndromes -
XX
PS Disclosure; Page 200-202; 216pp; English.
XX
CC The present sequence is that of cDNA encoding the protease domain
CC of human type II transmembrane serine protease 1 (MTSP1). The
CC invention relates to novel human type II transmembrane serine protein
CC 20 (MTSP20) (see ABP72374) and nucleic acids encoding it (see
CC ABZ58499). Also claimed are methods of inhibiting tumour initiation,
CC growth or progression by inhibiting MTSP20 activity, and of treating
CC or preventing a disease or disorder associated with undesired and/or
CC uncontrolled angiogenesis or neovascularisation, especially
CC undesired angiogenesis associated with solid neoplasms, vascular
CC malformations and cardiovascular disorders, chronic inflammatory
CC diseases, aberrant wound repairs, circulatory disorders, crest
CC syndromes, dermatological disorders and ocular disorders using an
CC inhibitor of MTSP20.
XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

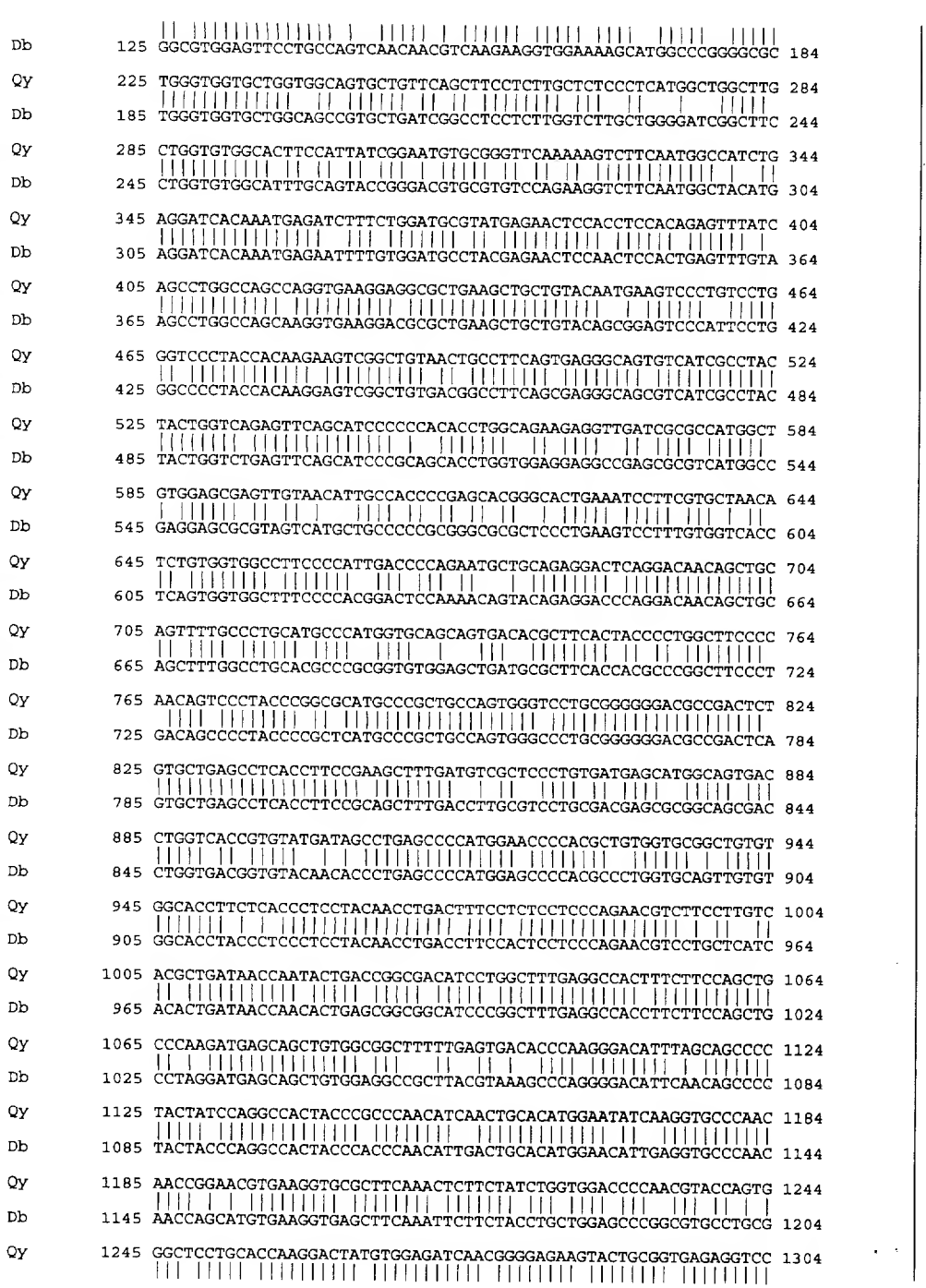


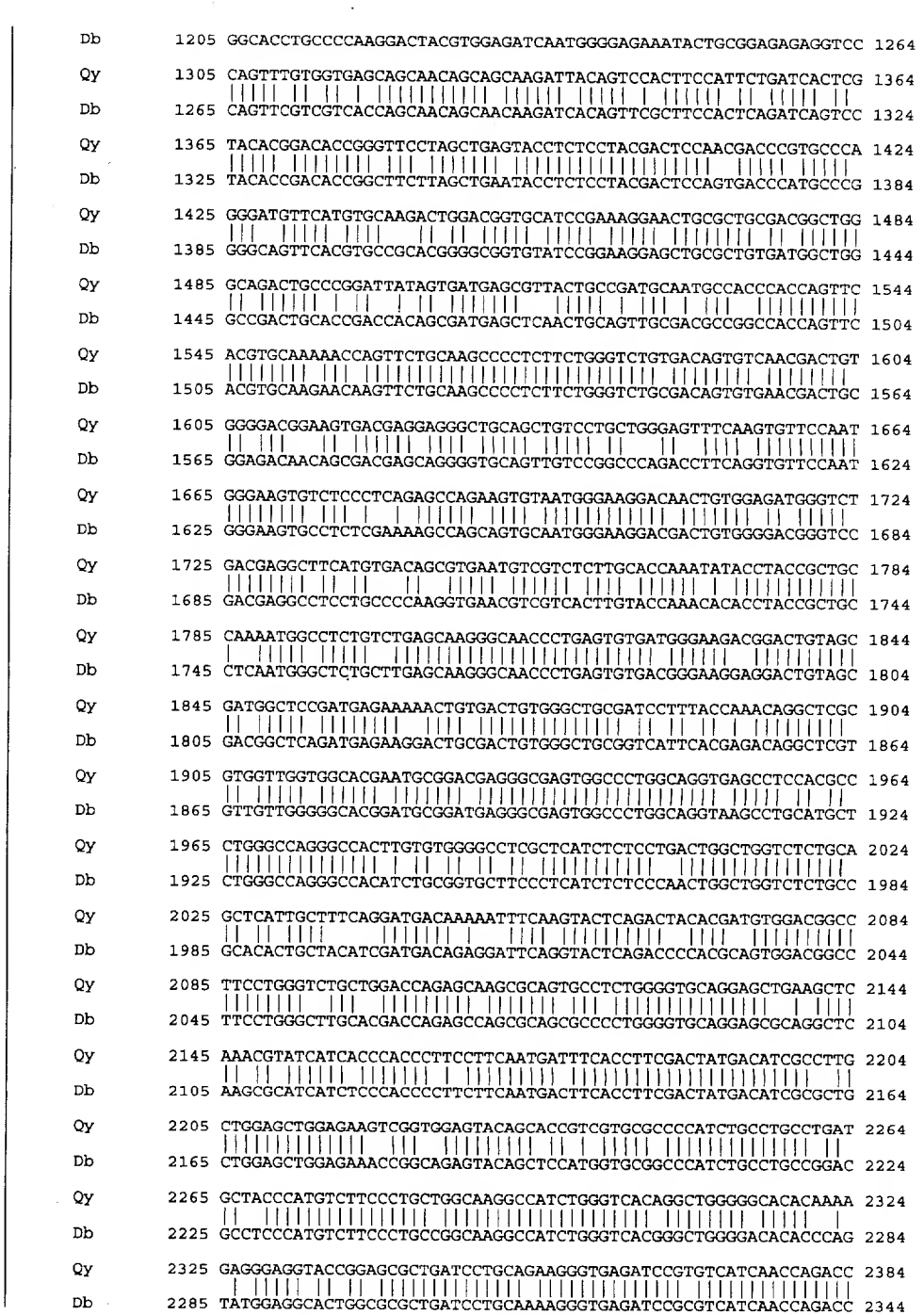


```
Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

```
RESULT 13
ABZ22450
ID   ABZ22450 standard; cDNA; 3147 BP.
XX
AC   ABZ22450;
XX
DT   24-MAR-2003  (first entry)
XX
DE   Human membrane-type serine protease MTSP1 encoding cDNA SEQ ID NO:1.
XX
KW   Human; membrane-type serine protease; enzyme; MTSP10; cytostatic;
KW   type-II membrane-type serine protease; neoplastic disease; tumour;
KW   MTSP1; matriptase; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             23..2590
FT                   /*tag=  a
FT                   /product= "membrane-type serine protease MTSP1"
FT                   /note= "also known as matriptase"
XX
PN   WO200292841-A2.
XX
PD   21-NOV-2002.
XX
PF   14-MAY-2002; 2002WO-US15332.
XX
PR   14-MAY-2001; 2001US-291001P.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Madison EL,  Yeh J;
XX
DR   WPI; 2003-129309/12.
DR   P-PSDB; ABP56619.
XX
PT   New polypeptides comprising the protease domain of a type-II
PT   membrane-type serine protease (MTSP10), or its muteins, useful for
PT   diagnosing neoplasms or malignancies, or for screening for MTSP10
PT   inhibitors for treating such diseases  -
XX
PS   Example 1; Page 177-181; 198pp; English.
XX
CC   The present invention describes a polypeptide comprising a purified
CC   single or two chain polypeptide, which comprises the protease domain of
CC   a type-II membrane-type serine protease (MTSP10) or its catalytically
CC   active portion, or a mutein of it, where up to 50 % of the amino acids
CC   are replaced with another amino acid, and the resulting polypeptide is a
CC   single chain or two chain polypeptide that has a catalytic activity of
CC   at least 1-10 % of the unmutated polypeptide. MTSP10 has cytostatic
CC   activity. The polypeptide containing the protease domain of the MTSP10
CC   is useful for detecting a neoplastic disease, and for diagnosing the
CC   presence of a pre-malignant lesion, a malignancy, or other pathologic
CC   condition in a subject, or monitoring tumour (e.g. breast, cervix,
CC   prostate, lung, ovary or colon tumour) progression and/or therapeutic
CC   effectiveness. An inhibitor of the polypeptide containing the protease
CC   domain of MTSP10 is useful for treating or preventing neoplastic disease
CC   in a mammal. An inhibitor of the activation cleavage of the zymogen
CC   form of the MTSP10 polypeptide is useful for inhibiting tumour
CC   initiation, growth or progression, or treating (pre-)malignant conditions
CC   of the e.g. breast, cervix, prostate, lung, ovary or colon. The present
CC   sequence encodes human MTSP1 (also known as matriptase), which is used
CC   in an example from the present invention.
XX
SQ   Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;
```

```
  Query Match             60.6%;  Score 1883.2;  DB 25;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
```



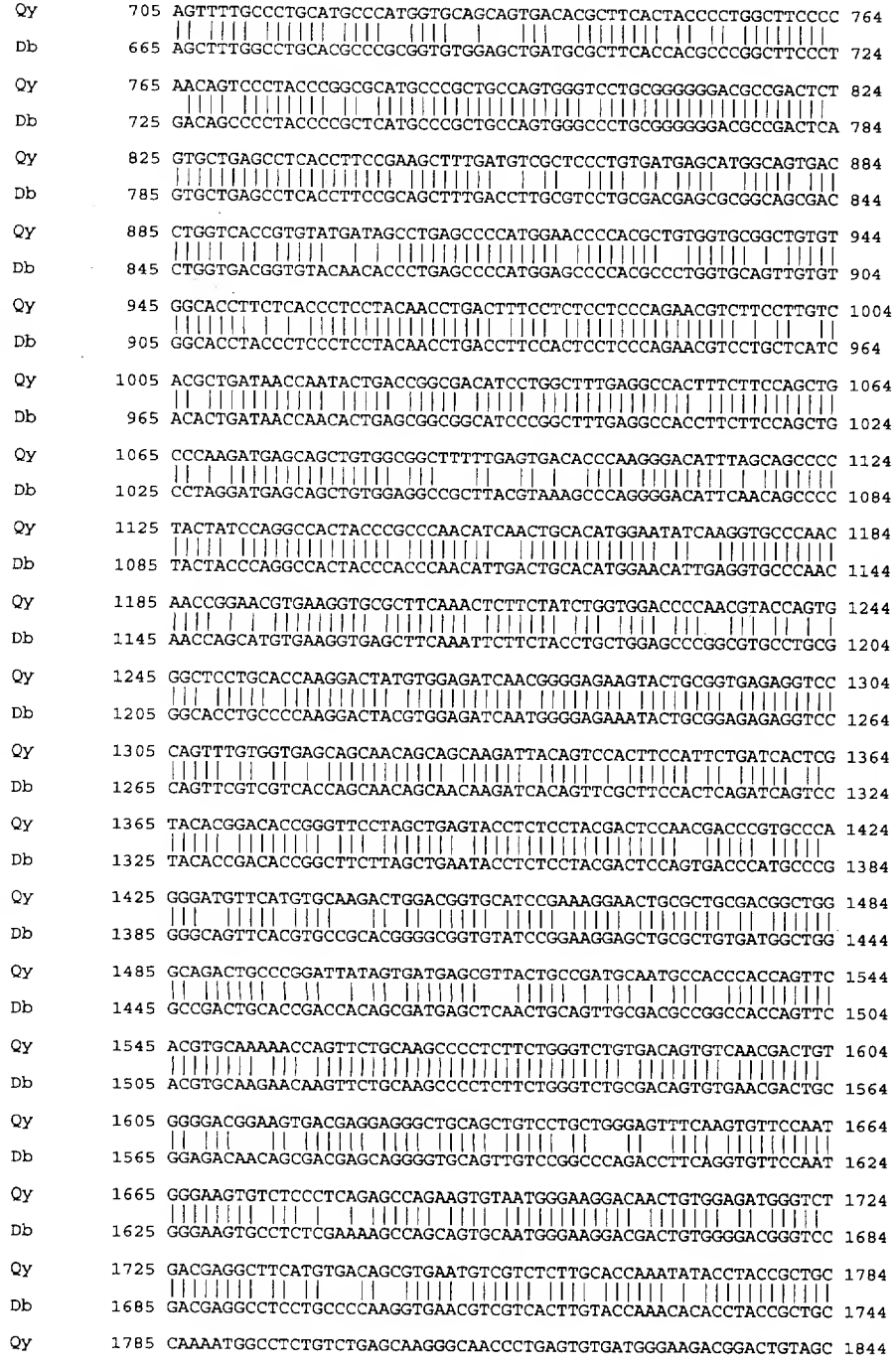


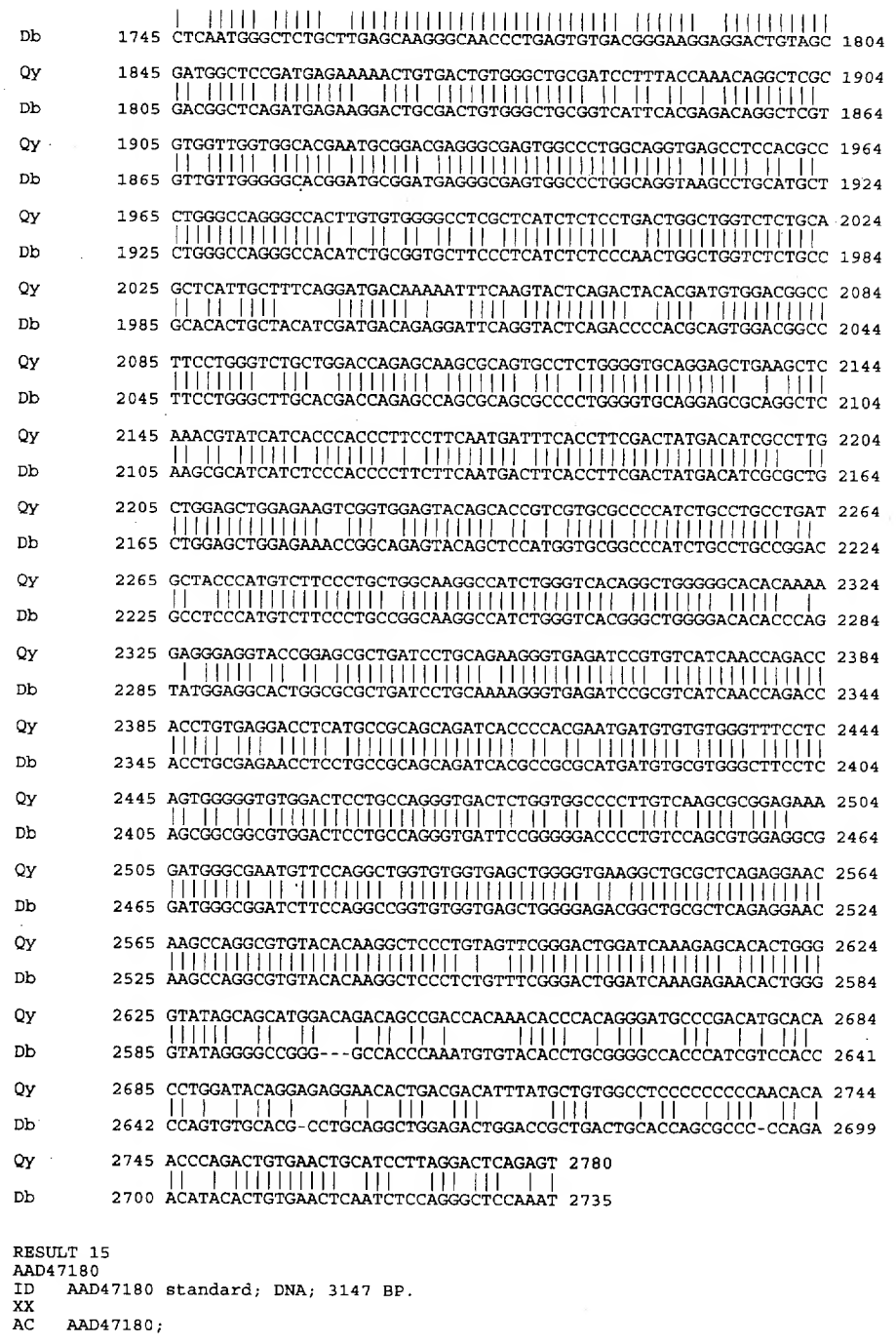
```
Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAA 2504
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 14
ABZ22451
ID ABZ22451 standard; cDNA; 3147 BP.
XX
AC ABZ22451;
XX
DT 24-MAR-2003 (first entry)
XX
DE Human MTSP1 protease domain encoding cDNA SEQ ID NO:3.
XX
KW Human; membrane-type serine protease; enzyme; MTSP10; cytostatic;
KW type-II membrane-type serine protease; neoplastic disease; tumour;
KW MTSP1; matriptase; gene; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1865..2590
FT   /*tag= a
FT   /partial
FT   /product= "MTSP1 protease domain"
FT   /note= "no start codon given"
XX
PN WO200292841-A2.
XX
PD 21-NOV-2002.
XX
PF 14-MAY-2002; 2002WO-US15332.
XX
PR 14-MAY-2001; 2001US-291001P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Yeh J;
XX
DR WPI; 2003-129309/12.
DR P-PSDB; ABP56620.
XX
PT New polypeptides comprising the protease domain of a type-II
PT membrane-type serine protease (MTSP10), or its muteins, useful for
PT diagnosing neoplasms or malignancies, or for screening for MTSP10
PT inhibitors for treating such diseases -
XX
PS Example 1; Page 183-185; 198pp; English.
XX
CC The present invention describes a polypeptide comprising a purified
CC single or two chain polypeptide, which comprises the protease domain of
CC a type-II membrane-type serine protease (MTSP10) or its catalytically
CC active portion, or a mutein of it, where up to 50 % of the amino acids
CC are replaced with another amino acid, and the resulting polypeptide is a
CC single chain or two chain polypeptide that has a catalytic activity of
CC at least 1-10 % of the unmutated polypeptide. MTSP10 has cytostatic
CC activity. The polypeptide containing the protease domain of the MTSP10
CC is useful for detecting a neoplastic disease, and for diagnosing the
CC presence of a pre-malignant lesion, a malignancy, or other pathologic
CC condition in a subject, or monitoring tumour (e.g. breast, cervix,
CC prostate, lung, ovary or colon tumour) progression and/or therapeutic
CC effectiveness. An inhibitor of the polypeptide containing the protease
CC domain of MTSP10 is useful for treating or preventing neoplastic disease
CC in a mammal. An inhibitor of the activation cleavage of the zymogen
CC form of the MTSP10 polypeptide is useful for inhibiting tumour
CC initiation, growth or progression, or treating (pre-)malignant conditions
CC of the e.g. breast, cervix, prostate, lung, ovary or colon. The present
CC sequence encodes the protease domain of human MTSP1 (also known as
CC matriptase), which is used in an example from the present invention.
XX
SQ Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

Query Match 60.6%; Score 1883.2; DB 25; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

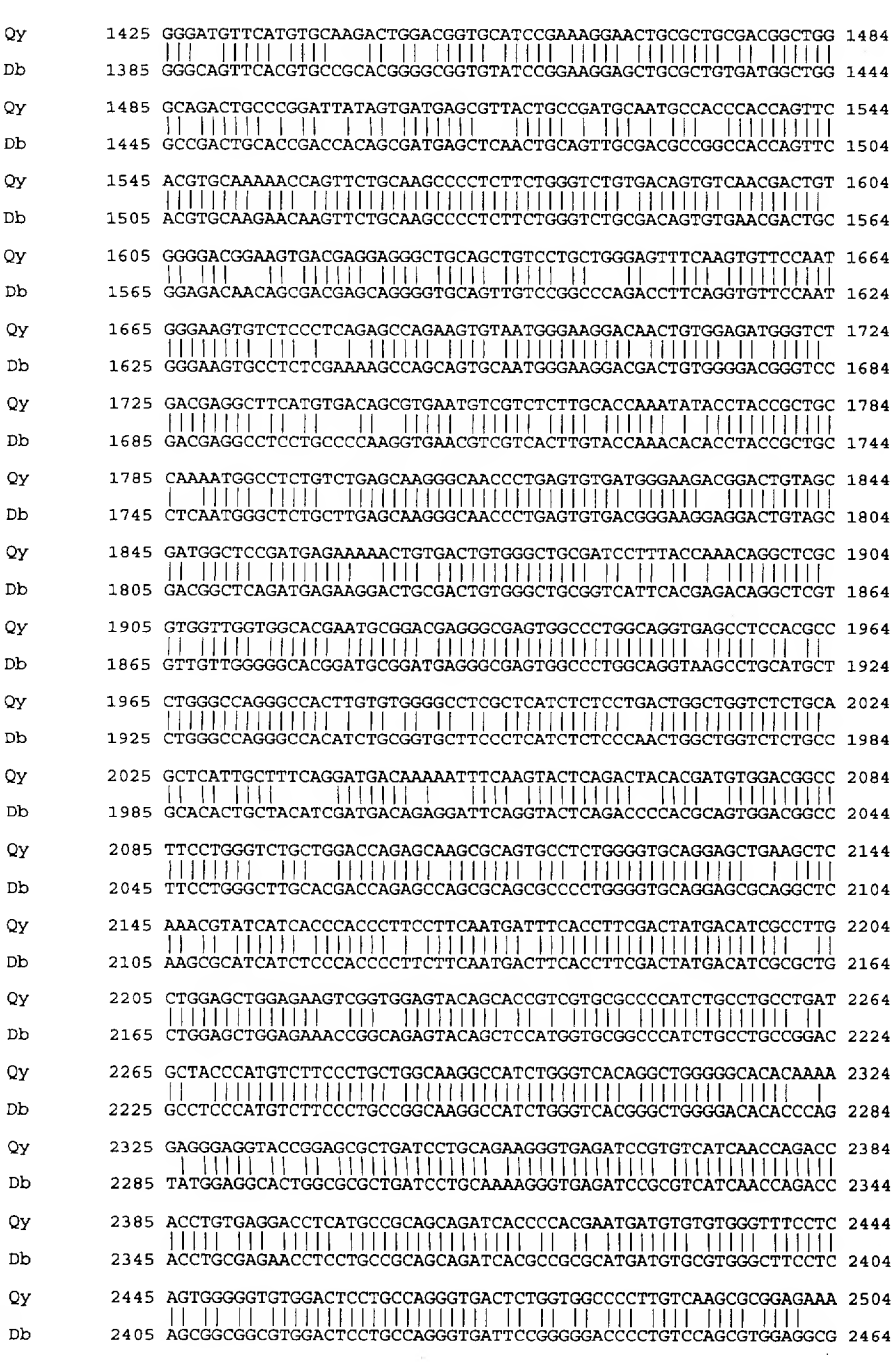
```
Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664
```





RESULT 15
AAD47180
ID AAD47180 standard; DNA; 3147 BP.
XX
AC AAD47180;

```
XX
DT   24-FEB-2003  (first entry)
XX
DE   Human membrane-type serine protease, MTSP1 DNA.
XX
KW   Human; serine protease 14; CVSP14; cancer; malignancy; breast; colon;
KW   gene therapy; cytostatic; membrane-type serine protease; enzyme; MTSP1;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             23..2590
FT                   /*tag= a
FT                   /product= "Human MTSP1 protein"
XX
PN   WO200277263-A2.
XX
PD   03-OCT-2002.
XX
PF   20-MAR-2002; 2002WO-US09039.
XX
PR   22-MAR-2001; 2001US-278166P.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Madison EL,  Yeh J;
XX
DR   WPI; 2003-018938/01.
DR   P-PSDB; AAE29791.
XX
PT   New purified CVSP14 polypeptide and encoding nucleic acid molecule,
PT   useful for diagnosing, preventing and/or treating disorders, such as
PT   cancers and malignancies of the breast, cervix, prostate, lung, ovary
PT   or colon  -
XX
PS   Disclosure; Page 167-171; 185pp; English.
XX
CC   The invention relates to transmembrane serine protease 14 (CVSP14), its
CC   nucleic acid sequence and the method based on them. The methods and
CC   compositions of the invention are useful for diagnosing, preventing
CC   and/or treating conditions associated with the aberrant expression or
CC   activity of the CVSP14 polypeptide, such as cancers and malignancies of
CC   the breast, cervix, prostate, lung, ovary or colon. The methods are
CC   also useful for identifying compounds that will modulate the protease
CC   activity of CVSP14 polypeptide, and monitoring tumour progression and/or
CC   therapeutic effectiveness. CVSP14 DNA used in gene therapy. The present
CC   sequence is human membrane-type serine protease, MTSP1 DNA.
XX
SQ   Sequence 3147 BP; 654 A; 952 C; 958 G; 583 T; 0 other;

  Query Match           60.6%;  Score 1883.2;  DB 25;  Length 3147;
  Best Local Similarity  81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy        45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
             ||  |||  |       |||||||| ||| |||||| |||||||| |||||||||| |  ||
Db         5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy       105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
             |||||||||||||||||||||||||||||||  |||||  |||||||||| |||||
Db        65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy       165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
             ||  |||||||||||||  |  |||||  |  |||||  |||||  ||||   |||||  |||||
Db       125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy       225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
             |||||||||||||  ||  ||||||  ||  ||  ||||||| |||  ||   |  |||||
Db       185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy       285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
             ||||||||||| ||  ||  ||  |||  |  |||||  ||  ||  ||  ||||||||||||  |  ||

Db       245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy       345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
             ||||||||||||||||  |||  |||||||  ||  |||||||||  ||||||  ||||||  |
Db       305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy       405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
             ||||||||||||  ||||||||||  ||||||||||||||||||||  |  ||||||  |||||
Db       365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy       465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
             ||  ||||||||||||  ||||||||||  ||  ||||||||  |||||||  |||||||||||
Db       425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy       525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
             |||||||  ||||||||||||||  |  |||||||  ||  |||  ||  ||||  ||||||
Db       485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy       585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
             |  ||||||  ||  ||  |   ||||  ||  ||  ||  ||  |  |||||  |||||  |||  |  ||
Db       545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy       645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
             ||  |||||||  |||||||  |||  |||  ||    |  ||||||||  ||||||||||||||||
Db       605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy       705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
             ||  ||||  ||||||  |||||  ||||  |  |||  ||||||||  ||  ||  ||||||||
Db       665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy       765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCT 824
             ||||  ||||||||  ||  ||||||||||||||||||  ||||||||||||||||||||
Db       725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGGACGCCGACTCA 784

Qy       825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
             ||||||||||||||||||||  ||||||||  |  ||  ||||  ||  ||||  |||||  |||
Db       785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy       885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
             |||||  ||  |||||  |  |  ||||||||||||||||  ||||||||  ||||||  |  |||||
Db       845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy       945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
             |||||||  |  ||||||||||||||||  ||||  ||||||||||||||||  |  ||  ||
Db       905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy      1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
             ||  ||||||||||  ||||  |||||  ||||||||||||  ||||||||||||||||||
Db       965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy      1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
             ||  |  ||||||||||||||  ||  ||  ||  |  ||||  ||||||||  |  |||||||
Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy      1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
             |||||  ||||||||||||||  |||||||  ||||||||||||  ||  ||||||||||
Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy      1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
             ||||  |  ||||||||||  |||||||  |||||||  |||  ||||  |||  |||  ||  |  |
Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy      1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
             |||  |||||  ||||||||||  ||||||||||  ||||||||  ||||||||  ||||||||||
Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy      1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
             |||||  ||  ||  |  ||||||||||  ||||||  |||||  |  ||||||  ||  |||||  ||
Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy      1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
             |||||  ||||||||  |||  |||||||  ||||||||||||||||||  |||||  ||||||
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
```



Search completed: February 21, 2004, 02:22:12
Job time : 812 secs

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:07:32 ; Search time 179 Seconds
(without alignments)
7658.860 Million cell updates/sec

Title: US-09-900-751-1
Perfect score: 3106
Sequence: 1 catggtagacggctgcccgg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

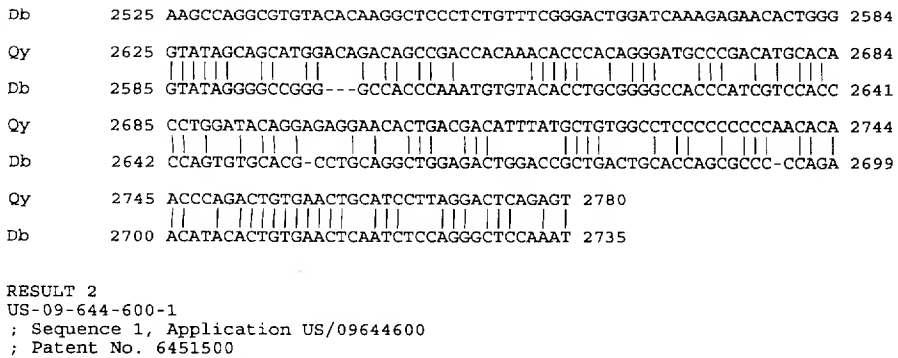
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1883.2	60.6	3147	2	US-09-027-337-1	Sequence 1, Appli
	2	1883.2	60.6	3147	4	US-09-644-600-1	Sequence 1, Appli
c	3	1883.2	60.6	3147	4	US-09-644-600-18	Sequence 18, Appl
	4	1530.8	49.3	2900	2	US-09-027-337-9	Sequence 9, Appli
	5	1530.8	49.3	2900	4	US-09-644-600-9	Sequence 9, Appli
	6	701.2	22.6	1553	4	US-09-280-116-10	Sequence 10, Appl
	7	306.8	9.9	434	4	US-09-702-705-1480	Sequence 1480, Ap
	8	306.8	9.9	434	4	US-09-736-457-1480	Sequence 1480, Ap
	9	173.2	5.6	796	4	US-09-280-116-107	Sequence 107, App
	10	113.6	3.7	2413	3	US-09-518-046-1	Sequence 1, Appli
	11	109.8	3.5	2544	3	US-09-518-046-3	Sequence 3, Appli
	12	101.6	3.3	1479	3	US-09-342-749-1	Sequence 1, Appli
	13	101.6	3.3	1479	4	US-09-691-840-1	Sequence 1, Appli
	14	101.2	3.3	1128	2	US-09-016-366A-20	Sequence 20, Appl
	15	101.2	3.3	1128	2	US-08-978-404B-15	Sequence 15, Appl
	16	100.6	3.2	1077	3	US-08-807-151-2	Sequence 2, Appli
	17	100.6	3.2	1077	4	US-09-478-957-2	Sequence 2, Appli
	18	99.6	3.2	1081	2	US-09-016-366A-22	Sequence 22, Appl
	19	99.6	3.2	1081	2	US-08-978-404B-17	Sequence 17, Appl
	20	98.4	3.2	2479	3	US-09-342-749-29	Sequence 29, Appl
	21	98.4	3.2	2479	4	US-09-691-840-29	Sequence 29, Appl
	22	98.2	3.2	735	3	US-09-079-970A-1	Sequence 1, Appli
	23	98.2	3.2	771	3	US-09-079-970A-4	Sequence 4, Appli
	24	98	3.2	1137	2	US-09-016-366A-18	Sequence 18, Appl
	25	98	3.2	1137	2	US-08-978-404B-13	Sequence 13, Appl
	26	94.2	3.0	2416	3	US-09-261-416-1	Sequence 1, Appli
	27	92.2	3.0	1605	2	US-09-000-846-1	Sequence 1, Appli
	28	91	2.9	980	4	US-09-023-942A-30	Sequence 30, Appl
	29	91	2.9	1110	4	US-09-386-653A-1	Sequence 1, Appli
	30	90.8	2.9	901	1	US-08-508-448C-9	Sequence 9, Appli
	31	90.8	2.9	1460	4	US-09-370-838-80	Sequence 80, Appl
	32	90.8	2.9	1462	4	US-09-370-838-55	Sequence 55, Appl
	33	90.8	2.9	1517	1	US-08-508-448C-15	Sequence 15, Appl
	34	90.8	2.9	2790	4	US-09-370-838-79	Sequence 79, Appl
	35	90.4	2.9	1783	3	US-09-510-738A-188	Sequence 188, App
	36	90.4	2.9	1783	4	US-09-861-966-188	Sequence 188, App
	37	90.4	2.9	2363	4	US-09-742-703-3	Sequence 3, Appli
	38	89.4	2.9	1154	2	US-09-016-366A-16	Sequence 16, Appl
	39	89.4	2.9	1154	2	US-08-978-404B-11	Sequence 11, Appl
	40	89.4	2.9	1212	4	US-09-620-312D-431	Sequence 431, App
	41	88.8	2.9	1130	4	US-09-386-653A-8	Sequence 8, Appli
	42	88.8	2.9	2038	3	US-09-008-271A-18	Sequence 18, Appl
	43	88.8	2.9	2079	4	US-09-656-002-1	Sequence 1, Appli
	44	88.6	2.9	1615	4	US-09-820-002-1	Sequence 1, Appli
	45	88.4	2.8	696	1	US-08-508-448C-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-09-027-337-1
; Sequence 1, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: 23..2589
; OTHER INFORMATION: cDNA sequence of TADG-15
US-09-027-337-1

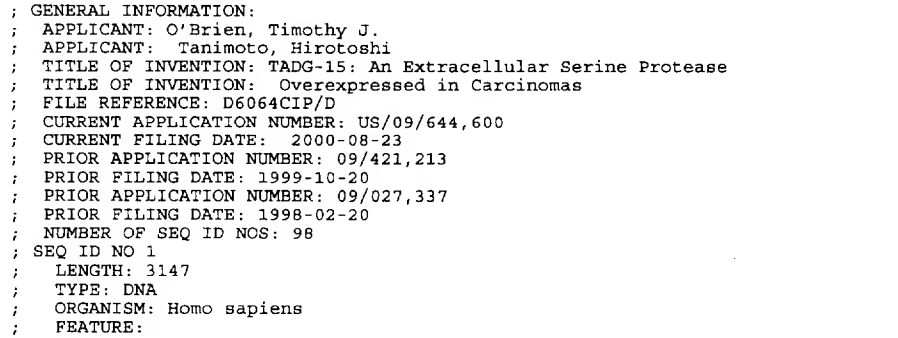
Query Match 60.6%; Score 1883.2; DB 2; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

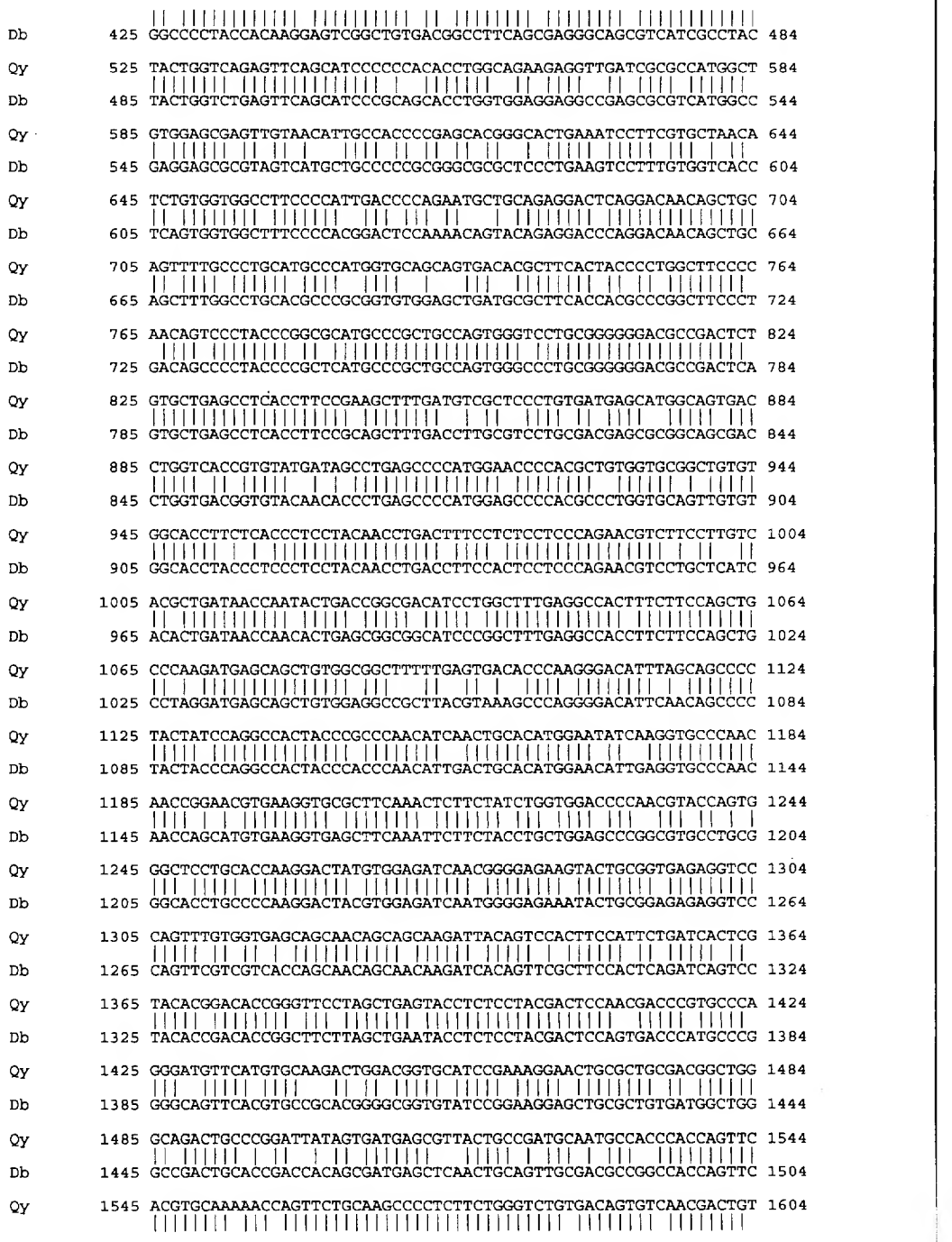
```
Qy          45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
               || ||| |     ||||||| ||| |||||| |||||||| |||||||||| |  ||
Db           5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy         105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
               ||||||||||||||||||||||||||||||||  |||||  |||||||||| |||||
Db          65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy         165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
               || |||||||||||||  |||||  | ||||||  |||||  ||||  |||||  |||||
Db         125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy         225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
               ||||||||||||  || ||||||  || ||  ||||||| |||  ||  |  |||||
Db         185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy         285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
               ||||||||||  ||  ||  ||| |  |||||  ||  ||  ||  ||||||||||||  |  ||
Db         245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy         345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
               ||||||||||||||||  |||  |||||||  ||  ||||||||||  ||||||  |||||  |
Db         305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364
```



RESULT 2
US-09-644-600-1
; Sequence 1, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-1

Query Match 60.6%; Score 1883.2; DB 4; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;



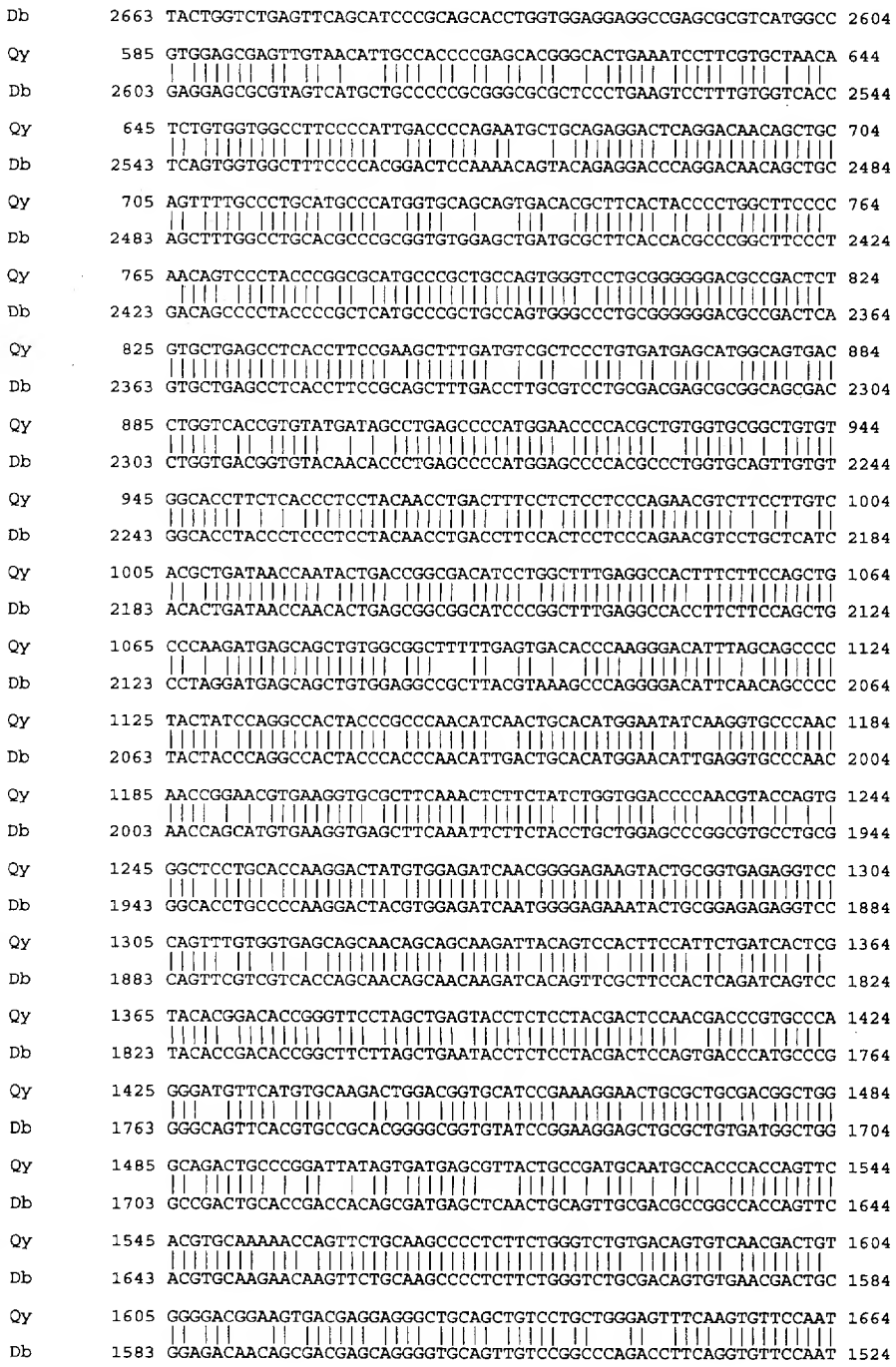


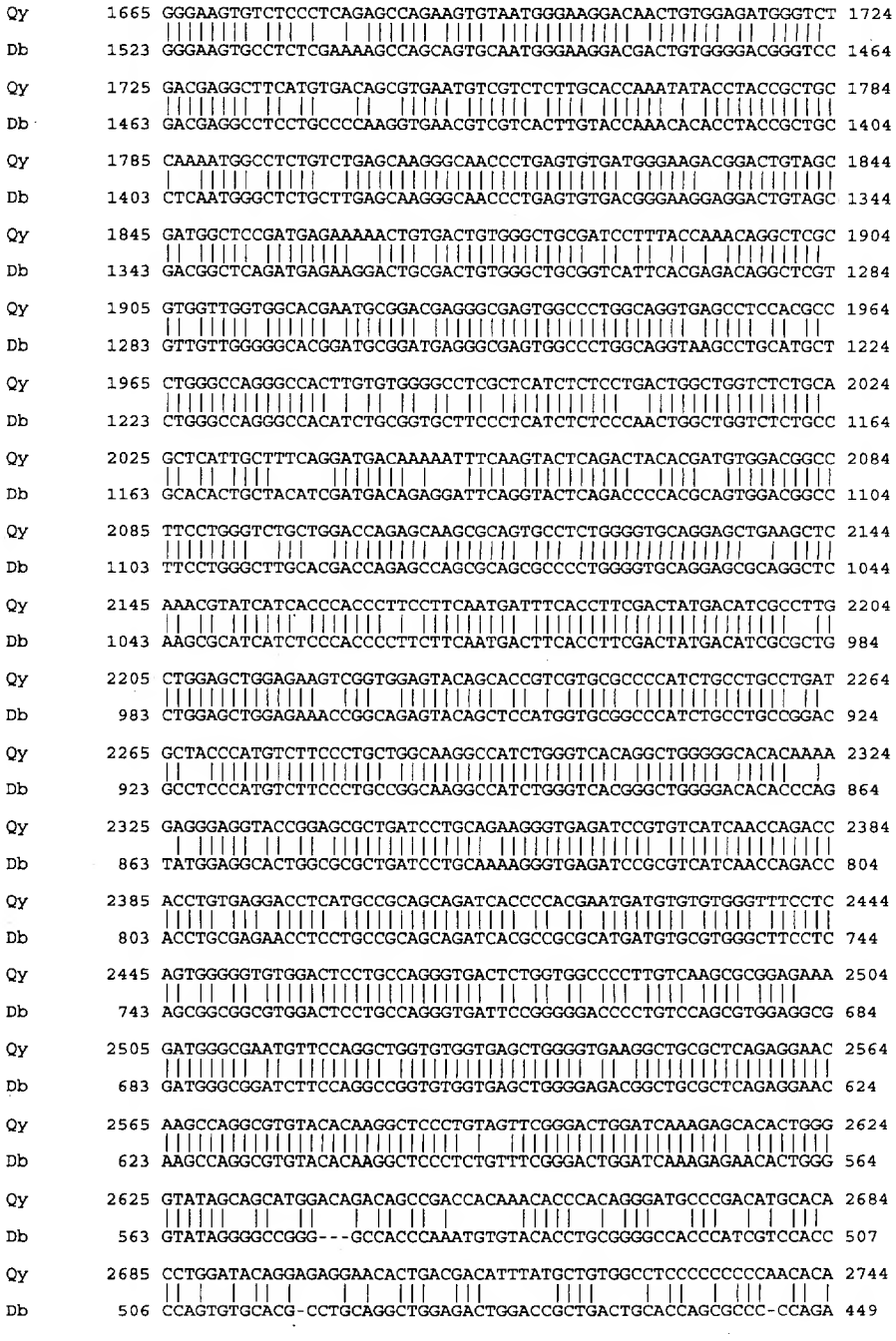
```
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664

Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724

Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784

Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844

Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904

Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964

Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024

Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084

Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144

Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204

Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264

Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324

Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384

Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444

Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504

Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 3
US-09-644-600-18/c
; Sequence 18, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 18
; LENGTH: 3147
; TYPE: RNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Antisense of TADG-15
US-09-644-600-18

Query Match 60.6%; Score 1883.2; DB 4; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

```
Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104

Db       3143 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 3084

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164

Db       3083 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 3024

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224

Db       3023 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 2964

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284

Db       2963 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 2904

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db       2903 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 2844

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404

Db       2843 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 2784

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464

Db       2783 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 2724

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524

Db       2723 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 2664

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
```


Db 2663 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 2604
Qy 585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db 2603 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 2544
Qy 645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db 2543 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 2484
Qy 705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db 2483 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 2424
Qy 765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db 2423 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 2364
Qy 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 2363 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 2304
Qy 885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db 2303 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 2244
Qy 945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db 2243 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 2184
Qy 1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db 2183 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 2124
Qy 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db 2123 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 2064
Qy 1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 2063 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 2004
Qy 1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db 2003 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1944
Qy 1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db 1943 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1884
Qy 1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db 1883 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1824
Qy 1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db 1823 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1764
Qy 1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db 1763 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1704
Qy 1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db 1703 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1644
Qy 1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1643 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1584
Qy 1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db 1583 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1524

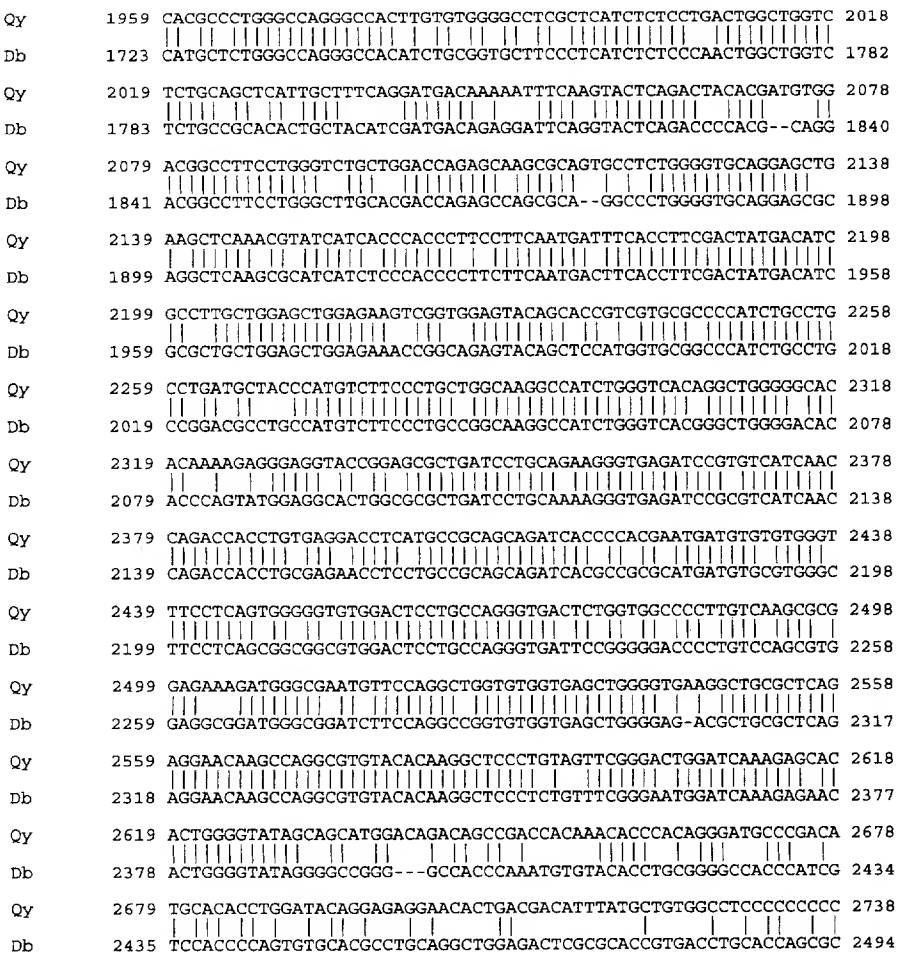

Qy 1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1523 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1464
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db 1463 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1404
Qy 1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db 1403 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1344
Qy 1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db 1343 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1284
Qy 1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db 1283 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1224
Qy 1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db 1223 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1164
Qy 2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db 1163 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 1104
Qy 2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 1103 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 1044
Qy 2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db 1043 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 984
Qy 2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db 983 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGAC 924
Qy 2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db 923 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 864
Qy 2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db 863 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 804
Qy 2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db 803 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 744
Qy 2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db 743 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 684
Qy 2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db 683 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 624
Qy 2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db 623 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 564
Qy 2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db 563 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 507
Qy 2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db 506 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 449

```
Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db        448 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 413
```

RESULT 4
US-09-027-337-9
; Sequence 9, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 9
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: SNC19 mRNA sequence (U20428)
US-09-027-337-9

Query Match 49.3%; Score 1530.8; DB 2; Length 2900;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 2011; Conservative 0; Mismatches 502; Indels 29; Gaps 15;

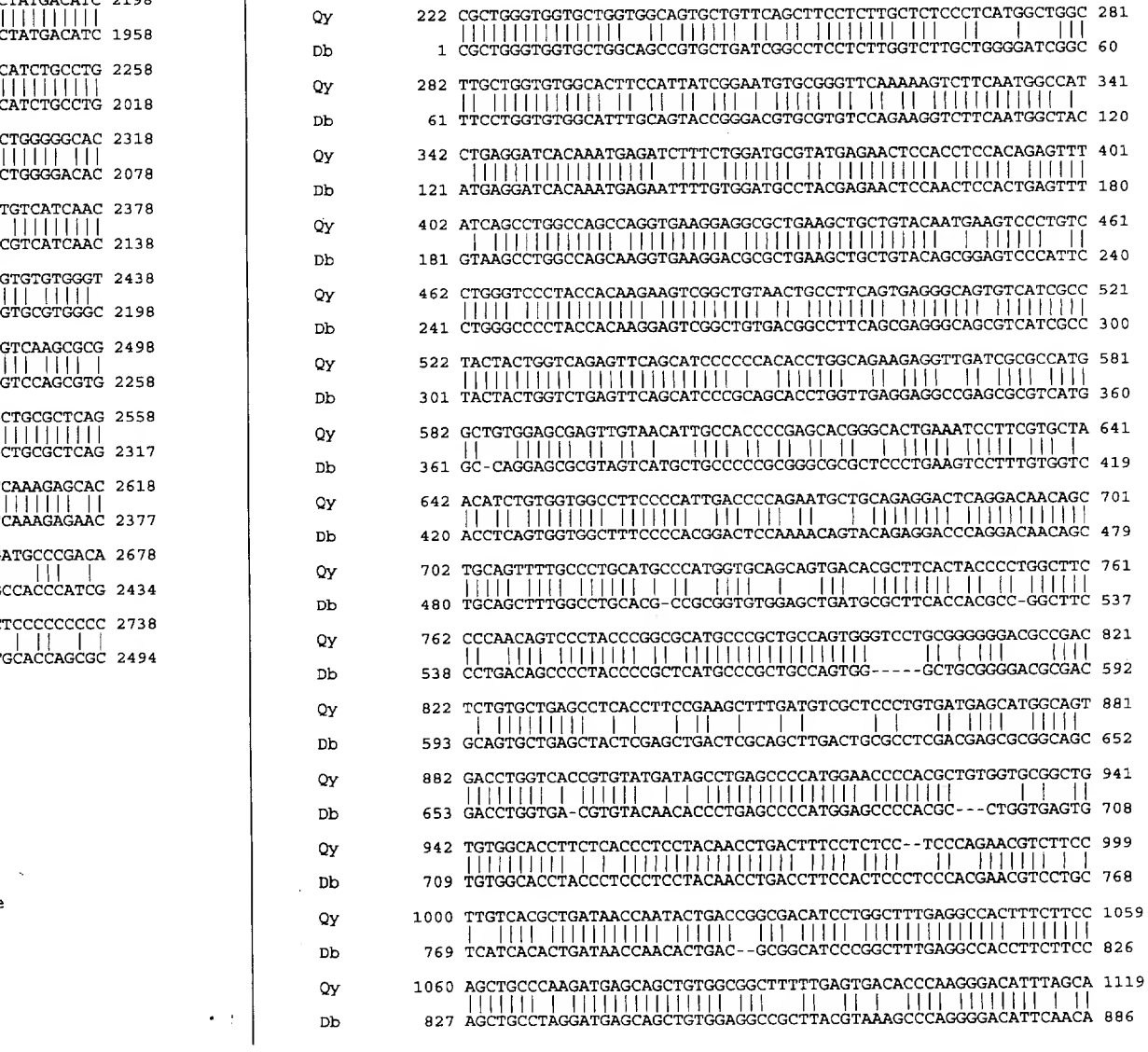
```
Qy        222 CGCTGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGC 281
Db          1 CGCTGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGC 60

Qy        282 TTGCTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCAT 341
Db         61 TTCCTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTAC 120

Qy        342 CTGAGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTT 401
Db        121 ATGAGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTT 180

Qy        402 ATCAGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTC 461
Db        181 GTAAGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTC 240

Qy        462 CTGGGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCC 521
Db        241 CTGGGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCC 300

Qy        522 TACTACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATG 581
Db        301 TACTACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTTGAGGAGGCCGAGCGCGTCATG 360

Qy        582 GCTGTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTA 641
Db        361 GC-CAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTC 419

Qy        642 ACATCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGC 701
Db        420 ACCTCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGC 479

Qy        702 TGCAGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTC 761
Db        480 TGCAGCTTTGGCCTGCACG-CCGCGGTGTGGAGCTGATGCGCTTCACCACGCC-GGCTTC 537

Qy        762 CCCAACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGAC 821
Db        538 CCTGACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGG-----GCTGCGGGGACGCGAC 592

Qy        822 TCTGTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGT 881
Db        593 GCAGTGCTGAGCTACTCGAGCTGACTCGCAGCTTGACTGCGCCTCGACGAGCGCGGCAGC 652

Qy        882 GACCTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTG 941
Db        653 GACCTGGTGA-CGTGTACAACACCCTGAGCCCCATGGAGCCCCACGC---CTGGTGAGTG 708

Qy        942 TGTGGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCC--TCCCAGAACGTCTTCC 999
Db        709 TGTGGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCCTCCCACGAACGTCCTGC 768

Qy       1000 TTGTCACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCC 1059
Db        769 TCATCACACTGATAACCAACACTGAC--GCGGCATCCCGGCTTTGAGGCCACCTTCTTCC 826

Qy       1060 AGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCA 1119
Db        827 AGCTGCCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACA 886

Qy       1120 GCCCCTACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGC 1179
Db        887 GCCCCTACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAAAATTGAGGTGC 946

Qy       1180 CCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTAC 1239
Db        947 CCAACAACCAGCATGTGAAGGTGCGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGC 1006

Qy       1240 CAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGA 1299
Db       1007 CTGCGGGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGA 1066

Qy       1300 GGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATC 1359
Db       1067 GGTCCCAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATC 1126

Qy       1360 ACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGT 1419
Db       1127 AGTCCTACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCAT 1186

Qy       1420 GCCCAGGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACG 1479
Db       1187 GCCCGGGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATG 1246

Qy       1480 GCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACC 1539
Db       1247 GCTGGGC-GACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACC 1305

Qy       1540 AGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACG 1599
Db       1306 AGTTCACGTGCAAGAGCAAGTTCTGCAAG---CTCTTCTGGGTCTGCGACAGTGTGAACG 1362

Qy       1600 ACTGTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTG-CTGGGAGTTTCAAGTGT 1658
Db       1363 AGTGCGGAGACAACAGCGACGAGCAGGGTTGCATTTGTCCGGACCCAGACCTTCAGGTGT 1422

Qy       1659 TCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGAT 1718
Db       1423 TCCAATGGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGAC 1482

Qy       1719 GGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTAC 1778
Db       1483 GGGTCCGACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTAC 1542

Qy       1779 CGCTGCCAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGAC 1838
Db       1543 CGCTGCCTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGAC 1602

Qy       1839 TGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAG 1898
Db       1603 TGTAGCGACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAG 1662

Qy       1899 GCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTC 1958
Db       1663 GCTCGTGTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTG 1722
```



RESULT 5
US-09-644-600-9
; Sequence 9, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20

; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 9
; LENGTH: 2900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: SNC-19; GeneBank Accession No. 6451500 #U20428
US-09-644-600-9

Query Match 49.3%; Score 1530.8; DB 4; Length 2900;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 2011; Conservative 0; Mismatches 502; Indels 29; Gaps 15;

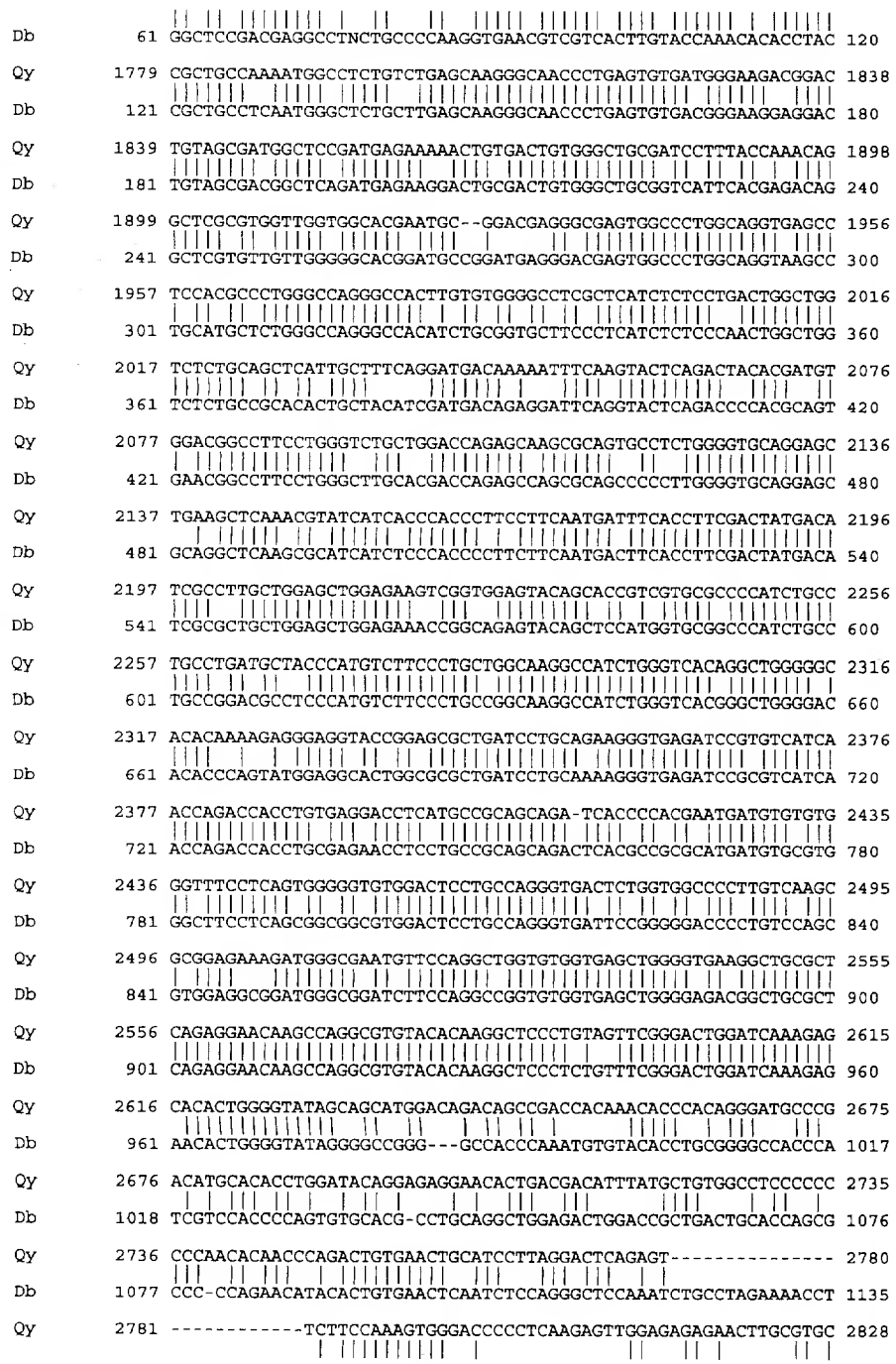


```
Qy       1120 GCCCCTACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGC 1179
Db        887 GCCCCTACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAAAATTGAGGTGC 946
Qy       1180 CCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTAC 1239
Db        947 CCAACAACCAGCATGTGAAGGTGCGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGC 1006
Qy       1240 CAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGA 1299
Db       1007 CTGCGGGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGA 1066
Qy       1300 GGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATC 1359
Db       1067 GGTCCCAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATC 1126
Qy       1360 ACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGT 1419
Db       1127 AGTCCTACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCAT 1186
Qy       1420 GCCCAGGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACG 1479
Db       1187 GCCCGGGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATG 1246
Qy       1480 GCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACC 1539
Db       1247 GCTGGGC-GACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACC 1305
Qy       1540 AGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACG 1599
Db       1306 AGTTCACGTGCAAGAGCAAGTTCTGCAAG---CTCTTCTGGGTCTGCGACAGTGTGAACG 1362
Qy       1600 ACTGTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTG-CTGGGAGTTTCAAGTGT 1658
Db       1363 AGTGCGGAGACAACAGCGACGAGCAGGGTTGCATTTGTCCGGACCCAGACCTTCAGGTGT 1422
Qy       1659 TCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGAT 1718
Db       1423 TCCAATGGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGAC 1482
Qy       1719 GGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTAC 1778
Db       1483 GGGTCCGACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTAC 1542
Qy       1779 CGCTGCCAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGAC 1838
Db       1543 CGCTGCCTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGAC 1602
Qy       1839 TGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAG 1898
Db       1603 TGTAGCGACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAG 1662
Qy       1899 GCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTC 1958
Db       1663 GCTCGTGTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTG 1722
Qy       1959 CACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTC 2018
Db       1723 CATGCTCTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTC 1782
Qy       2019 TCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGG 2078
Db       1783 TCTGCCGCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACG--CAGG 1840
Qy       2079 ACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTG 2138
Db       1841 ACGGCCTTCCTGGGCTTGCACGACCAGAGCCAGCGCA--GGCCCTGGGGTGCAGGAGCGC 1898
Qy       2139 AAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATC 2198
Db       1899 AGGCTCAAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATC 1958
Qy       2199 GCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTG 2258
Db       1959 GCGCTGCTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTG 2018
Qy       2259 CCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCAC 2318
Db       2019 CCGGACGCCTGCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACAC 2078
Qy       2319 ACAAAAGAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAAC 2378
Db       2079 ACCCAGTATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAAC 2138
Qy       2379 CAGACCACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGT 2438
Db       2139 CAGACCACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGC 2198
Qy       2439 TTCCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCG 2498
Db       2199 TTCCTCAGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTG 2258
Qy       2499 GAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAG 2558
Db       2259 GAGGCGGATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAG-ACGCTGCGCTCAG 2317
Qy       2559 AGGAACAAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCAC 2618
Db       2318 AGGAACAAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGAATGGATCAAAGAGAAC 2377
Qy       2619 ACTGGGGTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACA 2678
Db       2378 ACTGGGGTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCG 2434
Qy       2679 TGCACACCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCC 2738
Db       2435 TCCACCCCAGTGTGCACGCCTGCAGGCTGGAGACTCGCGCACCGTGACCTGCACCAGCGC 2494
Qy       2739 AACACAACCCAGACTGTGAACT 2760
Db       2495 CCCAGAACATACACTGTGAACT 2516
```

RESULT 6
US-09-280-116-10
; Sequence 10, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1553)
; OTHER INFORMATION: n = a, t, c, or g
US-09-280-116-10

Query Match 22.6%; Score 701.2; DB 4; Length 1553;
Best Local Similarity 71.1%; Pred. No. 1.2e-177;
Matches 1092; Conservative 0; Mismatches 351; Indels 92; Gaps 9;

```
Qy       1659 TCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGAT 1718
Db          1 TCCGATGNGAAGTGCCTNTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGAC 60
Qy       1719 GGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTAC 1778
```



RESULT 7
US-09-702-705-1480
; Sequence 1480, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1480
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-1480

Query Match 9.9%; Score 306.8; DB 4; Length 434;
Best Local Similarity 82.1%; Pred. No. 2e-72;
Matches 353; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy       1323 AACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGTTC 1382
             ||||||| |||||| ||||| | |||||| || ||||| || ||||| |||||||| |||
Db        241 AACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCCTACACCGACACCGGCTTC 300

Qy       1383 CTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAG 1442
             ||||||| ||||||||||||||||||||  ||||| ||||| |||  ||||| ||||
Db        301 TTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCGGGGCAGTTCACGTGCCGC 360

Qy       1443 ACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTAT 1502
             || || ||||| ||||| ||||| |||||||| || |||||||| |||||| | ||  |
Db        361 ACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGGGCCGACTGCACCGACCAC 420

Qy       1503 AGTGATGAGC 1512
             || |||||||
Db        421 AGCGATGAGC 430
```

RESULT 8
US-09-736-457-1480
; Sequence 1480, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1480
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-1480

Query Match 9.9%; Score 306.8; DB 4; Length 434;
Best Local Similarity 82.1%; Pred. No. 2e-72;
Matches 353; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy       1083 GGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACTAC 1142
             || |||   ||  || |  |||| |||||||| | |||||||||||| ||||||||||||
Db          1 GGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCCTACTACCCAGGCCACTAC 60

Qy       1143 CCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTG 1202
             || |||||||  |||||||||||||| ||  |||||||||||||||| | | |||||||||
Db         61 CCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAACAACCAGCATGTGAAGGTG 120

Qy       1203 CGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGGAC 1262
             |||||||| ||||||| ||| |||| |||  ||| || | |||| |||| ||||||||
Db        121 CGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCGGGCACCTGCCCCAAGGAC 180

Qy       1263 TATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGC 1322
             || |||||||||| |||||||| |||||||| |||||||||||||| || || | ||||
Db        181 TACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCCCAGTTCGTCGTCACCAGC 240

Qy       1323 AACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGTTC 1382
             ||||||| ||||||| ||||| | |||||| || ||||| || ||||| |||||||| |||
Db        241 AACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCCTACACCGACACCGGCTTC 300

Qy       1383 CTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAG 1442
             |||||||||||||||||||||||  ||||| ||||| |||  ||||| ||||
Db        301 TTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCGGGGCAGTTCACGTGCCGC 360

Qy       1443 ACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTAT 1502
             || || ||||| ||||| ||||| |||||||| || |||||||| |||||| | ||  |
Db        361 ACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGGGCCGACTGCACCGACCAC 420

Qy       1503 AGTGATGAGC 1512
             || |||||||
Db        421 AGCGATGAGC 430
```

RESULT 9
US-09-280-116-107
; Sequence 107, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 107
; LENGTH: 796
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(796)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-107

Query Match 5.6%; Score 173.2; DB 4; Length 796;
Best Local Similarity 57.7%; Pred. No. 1.4e-36;
Matches 395; Conservative 0; Mismatches 269; Indels 21; Gaps 4;

```
Qy       1744 GCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTGA 1803
             ||  | | |  ||  | ||   | ||  |||| ||  ||  | | |  |||  |||
Db         106 GCCAGGAAGGGGTGCCATGTGGGACATTCACCTTCCAGTGTGAGGACCGGAGCTGCGTGA 165

Qy       1804 GCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAAA 1863
             |||  ||||||  |||||||||||  | | ||||| || || |||||  ||||||  |
Db         166 AGAAGCCCAACCCGCAGTGTGATGGGCGGCCCGACTGCAGGGACGGCTCGGATGAGGAGC 225

Qy       1864 ACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATG 1923
             |||||||||||||| || |    |  |||         ||| | ||||||||  |
Db         226 ACTGTGACTGTGGCCTCCAGGGCCCCTCCA------GCCGCATTGTTGGTGGAGCTGTGT 279

Qy       1924 CGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTGT 1983
             |   ||||||  |||||||| |||||||  ||||||||  |  | |||  |   | |||  | |
Db         280 CCTCCGAGGGTGAGTGGCCATGGCAGGCCAGCCTCCAGGTTCGGGGTC---GACACATCT 336

Qy       1984 GTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATG 2043
             ||||||   | ||||||  ||     |||| ||  |  | ||  ||  || ||||| ||||| |
Db         337 GTGGGGGGGCCCTCATCGCTGACCGCTGGGTGATAACAGCTGCCCACTGCTTCCAGGAGG 396

Qy       2044 ACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACC 2103
             |||  |          |   |     | |||||||  |  ||||||||   | ||   |
Db         397 ACAGCA---------TGGCCTCCACGGTGCTGTGGACCGTGTTCCTGGGCAAGGTGTGGC 447

Qy       2104 AGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCCACC 2163
             |||  |  ||||  |     ||||  | |   |     | |||  |  |  ||  |  ||||
Db         448 AGAACTCGCGCTG---GCCTGGAGAGGTGTCCTTCAAGGTGAGCCGCCTGCTCCTGCACC 504

Qy       2164 CTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGTCGG 2223
             | | |  | | ||   || |   |||||  |||  | ||  |||||  ||||  ||   |  |||
Db         505 CGTACCACGAAGAGGACAGCCATGACTACGACGTGGCGCTGCTGCAGCTCGACCACCCGG 564

Qy       2224 TGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCCCTG 2283
             |||  |   |||  |||||||||| |||||||||| |       ||||  |||||
Db         565 TGGTGCGCTCGGCCGCCGTGCGCCCCGTCTGCCTGCCCGCGCGCTCCCACTTCTTCGAGC 624
```

```
Qy       2284 CTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAGCGC 2343
Db        625 CCGGCCTGCACTGCTGGATTACGGGCTGGGGCGCCTTGCGCGAGGGCGGCCCCATCAGCA 684

Qy       2344 TGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCATGC 2403
Db        685 ACGCTCTGCAGAAAGTGGATGTGCANTTGATCCCACAGGACCTGTGCAGCGAGGTCTATC 744

Qy       2404 CGCAGCAGATCACCCCACGAATGAT 2428
Db        745 GCTACCAGGTGACGCCACGCATGCT 769
```

RESULT 10
US-09-518-046-1
; Sequence 1, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 2413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
US-09-518-046-1

Query Match 3.7%; Score 113.6; DB 3; Length 2413;
Best Local Similarity 52.6%; Pred. No. 2.1e-20;
Matches 388; Conservative 0; Mismatches 314; Indels 36; Gaps 5;

```
Qy       1900 CTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCC 1959
Db        790 CACGCATCGTGGGTGGAAACATGTCCTTGCTCTCGCAGTGGCCCTGGCAGGCCAGCCTTC 849

Qy       1960 ACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCT 2019
Db        850 AGTTCCAGGGCTA---CCACCTGTGCGGGGGCTCTGTCATCACGCCCCTGTGGATCATCA 906

Qy       2020 CTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGA 2079
Db        907 CTGCTGCACACTGTGTTTATGACTTGTACCTCCCCAAGTCATGGACCATCCAGGTGGGTC 966

Qy       2080 CGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGA 2139
Db        967 TAGTTTCCCTG-------TTGGACAATCCAGCCCCATCCCACTTGGTGGAGAAG------ 1013

Qy       2140 AGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCG 2199
Db       1014 -----------ATTGTCTACCACAGCAAGTACAAGCCAAAGAGGCTGGGCAATGACATCG 1062

Qy       2200 CCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGC 2259
Db       1063 CCCTTATGAAGCTGGCCGGGCCACTCACGTTCAATGAAATGATCCAGCCTGTGTGCCTGC 1122

Qy       2260 CTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACA 2319
Db       1123 CCAACTCTGAAGAGAACTTCCCCGATGGAAAAGTGTGCTGGACGTCAGGATGGGGGGCCA 1182

Qy       2320 CAAAAGAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACC 2379
Db       1183 CAGAGGATGGAGGTGACGCCTCCCCTGTCCTGAACCACGCGGCCGTCCCTTTGATTTCCA 1242

Qy       2380 AGACCACCTGTGA------GGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTG 2433
Db       1243 ACAAGATCTGCAACCACAGGGACGTGTACGGTGGCATCATCTCCCCCTCCATGCTCTGCG 1302

Qy       2434 TGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAA 2493
Db       1303 CGGGCTACCTGACGGGTGGCGTGAACAGCTGCCAGGGGGACAGCGGGGGGCCCCTGGTGT 1362

Qy       2494 GCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCG 2553
Db       1363 GTCAAGAGAGGAGGCTGTGGAAGTT---AGTGGGAGCGACCAGCTTTGGCATCGGCTGCG 1419

Qy       2554 CTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAG 2613
Db       1420 CAGAGGTGAACAAGCCTGGGGTGTACACCCGTGTCACCTCCTTCCTGGACTGGATCCACG 1479

Qy       2614 AGCACACTGGGGTATAGC 2631
Db       1480 AGCAGATGGAGAGAGACC 1497

RESULT 11
US-09-518-046-3
; Sequence 3, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gene
US-09-518-046-3

Query Match 3.5%; Score 109.8; DB 3; Length 2544;
Best Local Similarity 54.6%; Pred. No. 2.3e-19;
Matches 266; Conservative 0; Mismatches 212; Indels 9; Gaps 2;

```
Qy   2151 ATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAG 2210
Db   1145 ATTGTCTACCACAGCAAGTACAAGCCAAAGAGGCTGGGCAATGACATCGCCCTTATGAAG 1204

Qy   2211 CTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACC 2270
Db   1205 CTGGCCGGGCCACTCACGTTCAATGAAATGATCCAGCCTGTGTGCCTGCCCAACTCTGAA 1264

Qy   2271 CATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGA 2330
Db   1265 GAGAACTTCCCCGATGGAAAAGTGTGCTGGACGTCAGGATGGGGGGCCACAGAGGATGGA 1324

Qy   2331 GGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGT 2390
Db   1325 GGTGACGCCTCCCCTGTCCTGAACCACGCGGCCGTCCCTTTGATTTCCAACAAGATCTGC 1384

Qy   2391 GA------GGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db   1385 AACCACAGGGACGTGTACGGTGGCATCATCTCCCCCTCCATGCTCTGCGCGGGCTACCTG 1444

Qy   2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db   1445 ACGGGTGGCGTGGACAGCTGCCAGGGGGACAGCGGGGGGCCCCTGGTGTGTCAAGAGAGG 1504
```
```

```
Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       1505 AGGCTGTGGAAGTT---AGTGGGAGCGACCAGCTTTGGCATCGGCTGCGCAGAGGTGAAC 1561

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       1562 AAGCCTGGGGTGTACACCCGTGTCACCTCCTTCCTGGACTGGATCCACGAGCAGATGGAG 1621

Qy       2625 GTATAGC 2631
Db       1622 AGAGACC 1628


RESULT 12
US-09-342-749-1
; Sequence 1, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
;  APPLICANT: Wong, Alexander K.C.
;  APPLICANT: Tavtigian, Sean V.
;  APPLICANT: Teng, DAvid H.-F.
;  APPLICANT: Myriad Genetics, Inc.
;  TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
;  FILE REFERENCE: 2318-202
;  CURRENT APPLICATION NUMBER: US/09/342,749
;  CURRENT FILING DATE: 1999-06-29
;  EARLIER APPLICATION NUMBER: US 60/091,044
;  EARLIER FILING DATE: 1998-06-29
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1479
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1476)
;   FEATURE:
;   NAME/KEY: conflict
;   LOCATION: (724)
;   OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
;   FEATURE:
;   NAME/KEY: conflict
;   LOCATION: (985)
;   OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
;   FEATURE:
;   NAME/KEY: conflict
;   LOCATION: (1347)
;   OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
;   FEATURE:
;   NAME/KEY: conflict
;   LOCATION: (1466)
;   OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
;   FEATURE:
;   NAME/KEY: conflict
;   LOCATION: (1471)
;   OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329.
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (478)
;   OTHER INFORMATION: This base can be G or A with G being the more
;   OTHER INFORMATION: common allele. The codon will change from Val to
;   OTHER INFORMATION: Met.
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (777)
;   OTHER INFORMATION: This base can be C or T with C being the more
;   OTHER INFORMATION: common allele.  The codon is unaffected with both
;   OTHER INFORMATION: alleles encoding Gly.
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (768)
;   OTHER INFORMATION: This base can be C or T with C being the more
;   OTHER INFORMATION: common allele.  This is a silent polymorphism.
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (834)
;   OTHER INFORMATION: This base can be C or T with C being the more
;   OTHER INFORMATION: common allele.  This is a silent polymorphism.
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (625)
;   OTHER INFORMATION: This base can be T or A with T being the more
;   OTHER INFORMATION: common allele.  The codon will change from Phe to
;   OTHER INFORMATION: Ile
US-09-342-749-1

  Query Match             3.3%;  Score 101.6;  DB 3;  Length 1479;
  Best Local Similarity   52.8%;  Pred. No. 2.7e-17;
  Matches  396;  Conservative    0;  Mismatches  324;  Indels   30;  Gaps    7;

Qy       1886 CTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTG 1945
Db        747 CTCAAGCCGCCAGAGCAGGATCGTGGGCGGCGAGAGCGCGCTCCCGGGGGCCTGGCCCTG 806

Qy       1946 GCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCC 2005
Db        807 GCAGGTCAGCCTGCACGTCCAGAAC---GTCCACGTGTGCGGAGGCTCCATCATCACCCC 863

Qy       2006 TGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGA 2065
Db        864 CGAGTGGATCGTGACAGCCGCCCACTGC------GTGGAAAAACCTCTTAACAATCCATG 917

Qy       2066 CTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGG 2125
Db        918 GCAT------TGGACGGCATTTGCGGGGATTTTGAGA-----CAATCTTTCATGTTCTAT 966

Qy       2126 GGTGCAGGAGCTGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTT 2185
Db        967 GGAGCCGGATACCAAG-TAGAAAAAGTGATTTCTCATCCAAATTATGACTCCAAGACCAA 1025

Qy       2186 CGACTATGACATCGCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCG 2245
Db        1026 GAACAATGACATTGCGCTGATGAAGCTGCAGAAGCCTCTGACTTTCAACGACCTAGTGAA 1085

Qy       2246 CCCCATCTGCCTGCCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCAC 2305
Db        1086 ACCAGTGTGTCTGCCCAACCCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTGGATTTC 1145

Qy       2306 AGGCTGGGGGCACACAAAAGAGGGAGGTA------CCGGAGCGCTGATCCTGCAGAAGGG 2359
Db        1146 CGGGTGGGGGGCCACCGAGGAGAAAGGGAAGACCTCAGAAGTGCTGAACGCTGCCAAGGT 1205

Qy       2360 TGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCATGCCGCAGCAGATCACCCC 2419
Db        1206 GCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACCTGATCACACC 1265

Qy       2420 ACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGG 2479
Db        1266 AGCCATGATCTGTGCCGGCTTCCTGCAGGGGAACGTCGATTCTTGCCAGGGTGACAGTGG 1325

Qy       2480 TGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTG 2539
Db        1326 AGGGCCTCTGGTCACTTCGAAGAACAATATCTG---GTGGCTGATAGGGGATACAAGCTG 1382

Qy       2540 GGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCG 2599
Db        1383 GGGTTCTGGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGGTATTCAC 1442

Qy       2600 GGACTGGATCAAAGAGCACACTGGGGTATA 2629
Db        1443 GGACTGGATTTATCGACAAATGAGGGCAGA 1472


RESULT 13
```

```
US-09-691-840-1
; Sequence 1, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
;  APPLICANT: Wong, Alexander K.C.
;  APPLICANT:  Tavtigian, Sean V.
;  APPLICANT:  Teng, DAvid H.-F.
;  APPLICANT:  Myriad Genetics, Inc.
;  TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
;  FILE REFERENCE: 2318-202
;  CURRENT APPLICATION NUMBER: US/09/691,840
;  CURRENT FILING DATE:  2000-10-18
;  PRIOR APPLICATION NUMBER: US/09/342,749
;  PRIOR FILING DATE: 1999-06-29
;  PRIOR APPLICATION NUMBER: US 60/091,044
;  PRIOR FILING DATE: 1998-06-29
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1479
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1476)
;   NAME/KEY: conflict
;   LOCATION: (724)
;   OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
;   NAME/KEY: conflict
;   LOCATION: (985)
;   OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
;   NAME/KEY: conflict
;   LOCATION: (1347)
;   OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
;   NAME/KEY: conflict
;   LOCATION: (1466)
;   OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
;   NAME/KEY: conflict
;   LOCATION: (1471)
;   OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329.
;   NAME/KEY: allele
;   LOCATION: (478)
;   OTHER INFORMATION: This base can be G or A with G being the more
;   OTHER INFORMATION: common allele. The codon will change from Val to
;   OTHER INFORMATION: Met.
;   NAME/KEY: allele
;   LOCATION: (777)
;   OTHER INFORMATION: This base can be C or T with C being the more
;   OTHER INFORMATION: common allele.  The codon is unaffected with both
;   OTHER INFORMATION: alleles encoding Gly.
;   NAME/KEY: allele
;   LOCATION: (768)
;   OTHER INFORMATION: This base can be C or T with C being the more
;   OTHER INFORMATION: common allele.  This is a silent polymorphism.
;   NAME/KEY: allele
;   LOCATION: (834)
;   OTHER INFORMATION: This base can be C or T with C being the more
;   OTHER INFORMATION: common allele.  This is a silent polymorphism.
;   NAME/KEY: allele
;   LOCATION: (625)
;   OTHER INFORMATION: This base can be T or A with T being the more
;   OTHER INFORMATION: common allele.  The codon will change from Phe to Ile
US-09-691-840-1

  Query Match             3.3%;  Score 101.6;  DB 4;  Length 1479;
  Best Local Similarity   52.8%;  Pred. No. 2.7e-17;
  Matches  396;  Conservative    0;  Mismatches  324;  Indels   30;  Gaps    7;

Qy       1886 CTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTG 1945
              ||  | |   |||     |  | || || |||  ||  |||  |  |||  |  ||||||||
Db        747 CTCAAGCCGCCAGAGCAGGATCGTGGGCGGCGAGAGCGCGCTCCCGGGGGCCTGGCCCTG 806

Qy       1946 GCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCC 2005
              ||||||  |||||  ||||  ||  |  |    |  ||||  ||||  ||  |  |||  |||||  |  ||
Db        807 GCAGGTCAGCCTGCACGTCCAGAAC---GTCCACGTGTGCGGAGGCTCCATCATCACCCC 863

Qy       2006 TGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGA 2065
              ||  |||  |  ||   |  ||  ||  ||  |||       |   ||  |||   |  |  ||   |   ||
Db        864 CGAGTGGATCGTGACAGCCGCCCACTGC------GTGGAAAAACCTCTTAACAATCCATG 917

Qy       2066 CTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGG 2125
              |       |||||||| ||    |||   |  ||          |||  |          |||
Db        918 GCAT------TGGACGGCATTTGCGGGGATTTTGAGA-----CAATCTTTCATGTTCTAT 966

Qy       2126 GGTGCAGGAGCTGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTT 2185
              ||  ||  |||     |||  |   ||      |  ||   |  ||  ||    |    |         |||
Db        967 GGAGCCGGATACCAAG-TAGAAAAAGTGATTTCTCATCCAAATTATGACTCCAAGACCAA 1025

Qy       2186 CGACTATGACATCGCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCG 2245
              ||  |||||||  ||   ||  ||||||  |||||  |   ||   |  ||  |   |  |||
Db       1026 GAACAATGACATTGCGCTGATGAAGCTGCAGAAGCCTCTGACTTTCAACGACCTAGTGAA 1085

Qy       2246 CCCCATCTGCCTGCCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCAC 2305
              ||  |  ||  ||||||  |   |   |     |  |  |   |  |   ||  |  ||||  |  |
Db       1086 ACCAGTGTGTCTGCCCAACCCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTGGATTTC 1145

Qy       2306 AGGCTGGGGGCACACAAAAGAGGGAGGTA------CCGGAGCGCTGATCCTGCAGAAGGG 2359
              ||  ||||||   |||  |  |||  |||  |       |  |  ||  |||||  |       ||||
Db       1146 CGGGTGGGGGGCCACCGAGGAGAAAGGGAAGACCTCAGAAGTGCTGAACGCTGCCAAGGT 1205

Qy       2360 TGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCATGCCGCAGCAGATCACCCC 2419
                  ||  ||  |   |  |   ||  |  |   |  ||    |  |  ||||||  ||
Db       1206 GCTTCTCATTGAGACACAGAGATGCAACAGCAGATATGTCTATGACAACCTGATCACACC 1265

Qy       2420 ACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGG 2479
              |   |||||  ||||  ||  |||||    |||   ||  ||  ||  ||||||||||||  |||
Db       1266 AGCCATGATCTGTGCCGGCTTCCTGCAGGGGAACGTCGATTCTTGCCAGGGTGACAGTGG 1325

Qy       2480 TGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTG 2539
              ||  ||  ||    |  ||  ||||  ||    |   ||  |  ||  |    |||||
Db       1326 AGGGCCTCTGGTCACTTCGAAGAACAATATCTG---GTGGCTGATAGGGGATACAAGCTG 1382

Qy       2540 GGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCG 2599
              ||||  |||||  ||  |    |||  |||||  ||||||   |   |    |||  |
Db       1383 GGGTTCTGGCTGTGCCAAAGCTTACAGACCAGGAGTGTACGGGAATGTGATGGTATTCAC 1442

Qy       2600 GGACTGGATCAAAGAGCACACTGGGGTATA 2629
              |||||||||  |    ||  |   |||  | |
Db       1443 GGACTGGATTTATCGACAAATGAGGGCAGA 1472


RESULT 14
US-09-016-366A-20
; Sequence 20, Application US/09016366A
; Patent No. 5955431
;  GENERAL INFORMATION:
;    APPLICANT:  Stevens, Richard L.
;    APPLICANT:  Huang, Chifu
;    TITLE OF INVENTION:  MAST CELL PROTEASE PEPTIDE
;    TITLE OF INVENTION:  INHIBITORS
;    NUMBER OF SEQUENCES:  65
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  U.S.A.
;      ZIP:  02210-2211
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/016,366A
;      FILING DATE:  January 30, 1998
```

```
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/037,090
;      FILING DATE:  05-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plumer, Elizabeth R.
;      REGISTRATION NUMBER:  36,637
;      REFERENCE/DOCKET NUMBER:  B0801/7093
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1128 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-016-366A-20

  Query Match             3.3%;  Score 101.2;  DB 2;  Length 1128;
  Best Local Similarity   53.1%;  Pred. No. 3.1e-17;
  Matches  394;  Conservative    0;  Mismatches  288;  Indels   60;  Gaps    6;

Qy       1903 GCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCC------ 1956
Db         87 GCATCGTTGGGGGTCAGGAGGCCCCCAGGAGCAAGTGGCCCTGGCAGGTGAGCCTGAGAG 146

Qy       1957 TCCACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGG 2016
Db        147 TCCACGGCCCATACTGGATGCACTTCTGCGGGGGCTCCCTCATCCACCCCCAGTGGGTGC 206

Qy       2017 TCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGT 2076
Db        207 TGACCGCAGCGCACTGCGTGGGACCGGACGTCAA----------------------GG 242

Qy       2077 GGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGC 2136
Db        243 ATCTGGCCGCCCTCAGGGTGCAACTGCGGGAGCAGCACCTCTACTACCAGGACCAGCTGC 302

Qy       2137 TGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACA 2196
Db        303 TGCCGGTCAGCAGGATCATCGTGCACCCACAGTTCTACACCGCCCAGATCGGAGCGGACA 362

Qy       2197 TCGCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCC 2256
Db        363 TCGCCCTGCTGGAGCTGGAGGAGCCGGTGAAGGTCTCCAGCCACGTCCACACGGTCACCC 422

Qy       2257 TGCCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGG-- 2314
Db        423 TGCCCCCTGCCTCAGAGACCTTCCCCCCGGGGATGCCGTGCTGGGTCACTGGCTGGGGCG 482

Qy       2315 ----GCACACAAAAGAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTG 2370
Db        483 ATGTGGACAATGATGAGCGCCTCCCACCGCCATTTCCTCTGAAGCAGGTGAAGGTCCCCA 542

Qy       2371 TCATCAACCAGACCACCTGTGA----------------GGACCTCATGCCGCAGCAGATC 2414
Db        543 TAATGGAAAACCACATTTGTGACGCAAAATACCACCTTGGCGCCTACACGGGAGACGACG 602

Qy       2415 ACCCCACGAATGATGTGT-----GTGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGG 2469
Db        603 TCCGCATCGTCCGTGACGACATGCTGTGTGCCGGGAACACCCGGAGGGACTCATGCCAGG 662

Qy       2470 GTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTG 2529
Db        663 GCGACTCCGGAGGGCCCCTGGTGTGCAAGGTGA---ATGGCACCTGGCTGCAGGCGGGCG 719

Qy       2530 TGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCC 2589
Db        720 TGGTCAGCTGGGGCGAGGGCTGTGCCCAGCCCAACCGGCCTGGCATCTACACCCGTGTCA 779
```

```
Qy       2590 CTGTAGTTCGGGACTGGATCAA 2611
Db        780 CCTACTACTTGGACTGGATCCA 801


RESULT 15
US-08-978-404B-15
; Sequence 15, Application US/08978404B
; Patent No. 5968782
;  GENERAL INFORMATION:
;    APPLICANT:  Stevens, Richard L.
;    TITLE OF INVENTION:  MAST CELL PROTEASE THAT CLEAVES
;    TITLE OF INVENTION:  FIBRINOGEN
;    NUMBER OF SEQUENCES:  74
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  U.S.A.
;      ZIP:  02210-2211
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,404B
;      FILING DATE:  25-NOV-97
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/032,354
;      FILING DATE:  04-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plumer, Elizabeth R.
;      REGISTRATION NUMBER:  36,637
;      REFERENCE/DOCKET NUMBER:  B0801/7090
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1128 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-978-404B-15

  Query Match             3.3%;  Score 101.2;  DB 2;  Length 1128;
  Best Local Similarity   53.1%;  Pred. No. 3.1e-17;
  Matches  394;  Conservative    0;  Mismatches  288;  Indels   60;  Gaps    6;

Qy       1903 GCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCC------ 1956
Db         87 GCATCGTTGGGGGTCAGGAGGCCCCCAGGAGCAAGTGGCCCTGGCAGGTGAGCCTGAGAG 146

Qy       1957 TCCACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGG 2016
Db        147 TCCACGGCCCATACTGGATGCACTTCTGCGGGGGCTCCCTCATCCACCCCCAGTGGGTGC 206

Qy       2017 TCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGT 2076
Db        207 TGACCGCAGCGCACTGCGTGGGACCGGACGTCAA----------------------GG 242

Qy       2077 GGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGC 2136
Db        243 ATCTGGCCGCCCTCAGGGTGCAACTGCGGGAGCAGCACCTCTACTACCAGGACCAGCTGC 302

Qy       2137 TGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACA 2196
Db        303 TGCCGGTCAGCAGGATCATCGTGCACCCACAGTTCTACACCGCCCAGATCGGAGCGGACA 362
```

```
Qy       2197 TCGCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCC 2256
Db        363 TCGCCCTGCTGGAGCTGGAGGAGCCGGTGAAGGTCTCCAGCCACGTCCACACGGTCACCC 422

Qy       2257 TGCCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGG-- 2314
Db        423 TGCCCCCTGCCTCAGAGACCTTCCCCCCGGGGATGCCGTGCTGGGTCACTGGCTGGGGCG 482

Qy       2315 ----GCACACAAAAGAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTG 2370
Db        483 ATGTGGACAATGATGAGCGCCTCCCACCGCCATTTCCTCTGAAGCAGGTGAAGGTCCCCA 542

Qy       2371 TCATCAACCAGACCACCTGTGA----------------GGACCTCATGCCGCAGCAGATC 2414
Db        543 TAATGGAAAACCACATTTGTGACGCAAAATACCACCTTGGCGCCTACACGGGAGACGACG 602

Qy       2415 ACCCCACGAATGATGTGT-----GTGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGG 2469
Db        603 TCCGCATCGTCCGTGACGACATGCTGTGTGCCGGGAACACCCGGAGGGACTCATGCCAGG 662

Qy       2470 GTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTG 2529
Db        663 GCGACTCCGGAGGGCCCCTGGTGTGCAAGGTGA---ATGGCACCTGGCTGCAGGCGGGCG 719

Qy       2530 TGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACACAAGGCTCC 2589
Db        720 TGGTCAGCTGGGGCGAGGGCTGTGCCCAGCCCAACCGGCCTGGCATCTACACCCGTGTCA 779

Qy       2590 CTGTAGTTCGGGACTGGATCAA 2611
Db        780 CCTACTACTTGGACTGGATCCA 801
```

Search completed: February 21, 2004, 07:15:18
Job time : 193 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 05:28:33 ; Search time 1052 Seconds
(without alignments)
10875.821 Million cell updates/sec

Title: US-09-900-751-1
Perfect score: 3106
Sequence: 1 catggtagacggctgcccgg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3106	100.0	3106	9	US-09-900-751-1	Sequence 1, Appli
2	1889.8	60.8	3149	12	US-10-295-027-968	Sequence 968, App
3	1883.2	60.6	3147	11	US-09-776-191-1	Sequence 1, Appli
4	1883.2	60.6	3147	11	US-09-776-191-49	Sequence 49, Appl
5	1883.2	60.6	3147	12	US-10-147-211A-1	Sequence 1, Appli
6	1883.2	60.6	3147	12	US-10-147-211A-3	Sequence 3, Appli
7	1883.2	60.6	3147	12	US-10-156-214A-1	Sequence 1, Appli
8	1883.2	60.6	3147	13	US-10-190-030B-1	Sequence 1, Appli
9	1883.2	60.6	3147	13	US-10-190-030B-3	Sequence 3, Appli
10	1883.2	60.6	3147	13	US-10-267-219-1	Sequence 1, Appli
11	1883.2	60.6	3147	13	US-10-267-219-3	Sequence 3, Appli
12	1883.2	60.6	3147	13	US-10-302-840A-1	Sequence 1, Appli
13	1883.2	60.6	3147	13	US-10-302-840A-3	Sequence 3, Appli
14	1883.2	60.6	3147	13	US-10-112-221A-1	Sequence 1, Appli
15	1883.2	60.6	3147	13	US-10-112-221A-3	Sequence 3, Appli
16	1883.2	60.6	3147	13	US-10-104-271-1	Sequence 1, Appli
17	1883.2	60.6	3147	13	US-10-104-271-3	Sequence 3, Appli
18	1883.2	60.6	3147	15	US-10-099-700A-1	Sequence 1, Appli
19	1883.2	60.6	3147	15	US-10-099-700A-3	Sequence 3, Appli
20	1559.2	50.2	2756	9	US-09-925-301-351	Sequence 351, App
21	1532.4	49.3	3112	15	US-10-097-340-311	Sequence 311, App
22	556.4	17.9	726	15	US-10-092-004A-1	Sequence 1, Appli
23	328	10.6	472	11	US-09-918-995-35415	Sequence 35415, A
24	306.8	9.9	434	10	US-09-736-457-1480	Sequence 1480, Ap
25	306.8	9.9	434	10	US-09-902-941-1480	Sequence 1480, Ap
26	306.8	9.9	434	10	US-09-849-626-1480	Sequence 1480, Ap
27	306.8	9.9	434	13	US-10-113-872-1480	Sequence 1480, Ap
28	306.8	9.9	434	15	US-10-017-754-1480	Sequence 1480, Ap
29	302.6	9.7	429	10	US-09-867-701-5310	Sequence 5310, Ap
30	265	8.5	591	9	US-09-815-343-1503	Sequence 1503, Ap
31	246.8	7.9	393	10	US-09-960-352-5954	Sequence 5954, Ap
32	240.8	7.8	2672	11	US-09-776-191-9	Sequence 9, Appli
33	240.8	7.8	2672	12	US-10-156-214A-9	Sequence 9, Appli
34	240.8	7.8	3104	11	US-09-776-191-7	Sequence 7, Appli
35	240.8	7.8	3104	12	US-10-156-214A-7	Sequence 7, Appli
36	239.2	7.7	1327	10	US-09-978-295A-170	Sequence 170, App
37	239.2	7.7	1327	10	US-09-978-697-170	Sequence 170, App
38	239.2	7.7	1327	10	US-09-978-192A-170	Sequence 170, App
39	239.2	7.7	1327	10	US-09-999-832A-170	Sequence 170, App
40	239.2	7.7	1327	11	US-09-978-189-170	Sequence 170, App
41	239.2	7.7	1327	11	US-09-978-608A-170	Sequence 170, App
42	239.2	7.7	1327	11	US-09-978-585A-170	Sequence 170, App
43	239.2	7.7	1327	11	US-09-978-191A-170	Sequence 170, App
44	239.2	7.7	1327	11	US-09-978-403A-170	Sequence 170, App
45	239.2	7.7	1327	11	US-09-978-564A-170	Sequence 170, App

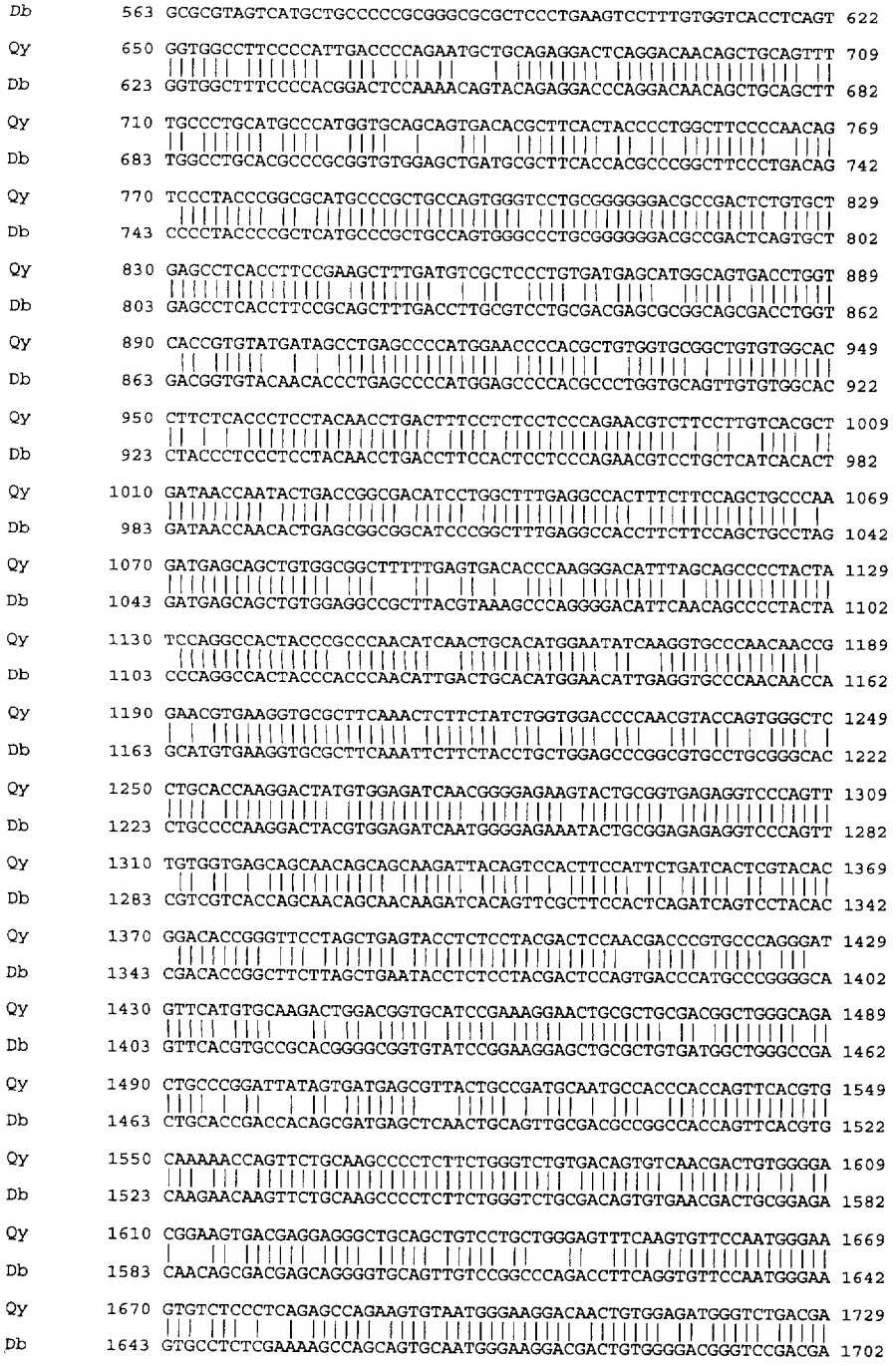
ALIGNMENTS

```
RESULT 1
US-09-900-751-1
; Sequence 1, Application US/09900751
; Patent No. US20020026653A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Keith D
;  APPLICANT:  Leviten, Michael W.
;  TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE
;  TITLE OF INVENTION:  PROTEASE GENE DISRUPTIONS
;  FILE REFERENCE: R-386
;  CURRENT APPLICATION NUMBER: US/09/900,751
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US 60/217,449
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: US 60/223,170
;  PRIOR FILING DATE: 2000-08-07
;  PRIOR APPLICATION NUMBER: US 60/223,460
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3106
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-900-751-1

  Query Match             100.0%;  Score 3106;  DB 9;  Length 3106;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3106;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
```

```
Qy      121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
Db      121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180

Qy      181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
Db      181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240

Qy      241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
Db      241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300

Qy      301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
Db      301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360

Qy      361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420
Db      361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420

Qy      421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480
Db      421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480

Qy      481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540
Db      481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540

Qy      541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600
Db      541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600

Qy      601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660
Db      601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660

Qy      661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720
Db      661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720

Qy      721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780
Db      721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780

Qy      781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840
Db      781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840

Qy      841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900
Db      841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900

Qy      901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960
Db      901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960

Qy      961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020
Db      961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020

Qy     1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080
Db     1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080

Qy     1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140
Db     1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140

Qy     1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
Db     1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200

Qy     1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
Db     1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260

Qy     1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
Db     1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320

Qy     1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
Db     1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380

Qy     1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
Db     1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440

Qy     1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
Db     1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500

Qy     1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
Db     1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560

Qy     1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
Db     1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620

Qy     1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
Db     1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680

Qy     1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
Db     1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740

Qy     1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
Db     1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800

Qy     1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
Db     1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860

Qy     1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
Db     1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920

Qy     1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
Db     1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980

Qy     1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
Db     1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040

Qy     2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100
Db     2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100

Qy     2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160
Db     2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160

Qy     2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220
Db     2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220

Qy     2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280
Db     2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280

Qy     2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
```

```
Db       2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340

Qy       2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400

Qy       2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460

Qy       2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520

Qy       2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580

Qy       2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640

Qy       2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700

Qy       2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2760

Qy       2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820

Qy       2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880

Qy       2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940

Qy       2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000

Qy       3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060

Qy       3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
             ||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
```

```
RESULT 2
US-10-295-027-968
; Sequence 968, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 968
;   LENGTH: 3149
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-295-027-968

  Query Match            60.8%;  Score 1889.8;  DB 12;  Length 3149;
  Best Local Similarity  81.4%;  Pred. No. 0;
  Matches 2222;  Conservative    0;  Mismatches  504;  Indels    5;  Gaps    3;

Qy         50 GACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTT 109
              | || |    |||||||| ||| ||||||| |||||||| |||||||||| | |||||||
Db         23 GGCCTCGGGGACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAGGACTT 82

Qy        110 CGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGT 169
              |||||||||||||||||||||||||||||  |||||  |||||||||| ||||| || ||
Db         83 CGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAAGGCGT 142

Qy        170 GGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGT 229
              |||||||||||  |||||  |||||| |||||  ||||  |||||  ||||||||||
Db        143 GGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGCTGGGT 202

Qy        230 GGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGT 289
              |||||||  || |||||| || || |||||||| |||  ||   |  ||||| |||||
Db        203 GGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTCCTGGT 262

Qy        290 GTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGAT 349
              ||||||  || ||  || ||| | |||||  || ||  || ||||||||||| |  |||||||
Db        263 GTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATGAGGAT 322

Qy        350 CACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCT 409
              ||||||||||  ||| |||||||  || ||||||||||  ||||||  ||||||  | |||||
Db        323 CACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTAAGCCT 382

Qy        410 GGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCC 469
              |||||||  ||||||||||  ||||||||||||||||||||  | ||||||  |||||||  ||
Db        383 GGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTGGGCCC 442

Qy        470 CTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTG 529
              ||||||||||  ||||||||||  ||  ||||||||  ||||||||||||||||||||||
Db        443 CTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTACTACTG 502

Qy        530 GTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGA 589
              |||  ||||||||||||||  |  |||||||  ||  ||||  ||  |||| ||||||  | |||
Db        503 GTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCCGAGGA 562

Qy        590 GCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGT 649
              |||  ||  ||  |   ||||  ||  ||  ||  ||  |  |||||  |||||  |||  |  ||  ||  ||
```


Db 563 GCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACCTCAGT 622
Qy 650 GGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTT 709
Db 623 GGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGCAGCTT 682
Qy 710 TGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAG 769
Db 683 TGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCTGACAG 742
Qy 770 TCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCT 829
Db 743 CCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCAGTGCT 802
Qy 830 GAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGT 889
Db 803 GAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGACCTGGT 862
Qy 890 CACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCAC 949
Db 863 GACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGTGGCAC 922
Qy 950 CTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCT 1009
Db 923 CTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATCACACT 982
Qy 1010 GATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAA 1069
Db 983 GATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTGCCTAG 1042
Qy 1070 GATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTA 1129
Db 1043 GATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCCTACTA 1102
Qy 1130 TCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCG 1189
Db 1103 CCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAACAACCA 1162
Qy 1190 GAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTC 1249
Db 1163 GCATGTGAAGGTGCGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCGGGCAC 1222
Qy 1250 CTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTT 1309
Db 1223 CTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCCCAGTT 1282
Qy 1310 TGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACAC 1369
Db 1283 CGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCCTACAC 1342
Qy 1370 GGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGAT 1429
Db 1343 CGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCGGGGCA 1402
Qy 1430 GTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGA 1489
Db 1403 GTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGGGCCGA 1462
Qy 1490 CTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTG 1549
Db 1463 CTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTCACGTG 1522
Qy 1550 CAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGA 1609
Db 1523 CAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGCGGAGA 1582
Qy 1610 CGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAA 1669
Db 1583 CAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAATGGGAA 1642
Qy 1670 GTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGA 1729
Db 1643 GTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCCGACGA 1702
Qy 1730 GGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAA 1789
Db 1703 GGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGCCTCAA 1762
Qy 1790 TGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGG 1849
Db 1763 TGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGCGACGG 1822
Qy 1850 CTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGT 1909
Db 1823 CTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGTGTTGT 1882
Qy 1910 TGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGG 1969
Db 1883 TGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCTCTGGG 1942
Qy 1970 CCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCA 2029
Db 1943 CCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCCGCACA 2002
Qy 2030 TTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCT 2089
Db 2003 CTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCCTTCCT 2062
Qy 2090 GGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACG 2149
Db 2063 GGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTCAAGCG 2122
Qy 2150 TATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGA 2209
Db 2123 CATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTGCTGGA 2182
Qy 2210 GCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTAC 2269
Db 2183 GCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGACGCCTC 2242
Qy 2270 CCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGG 2329
Db 2243 CCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAGTATGG 2302
Qy 2330 AGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTG 2389
Db 2303 AGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACCACCTG 2362
Qy 2390 TGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGG 2449
Db 2363 CGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTCAGCGG 2422
Qy 2450 GGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGG 2509
Db 2423 CGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCGGATGG 2482
Qy 2510 GCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCC 2569
Db 2483 GCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAACAAGCC 2542
Qy 2570 AGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATA 2629
Db 2543 AGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGGGTATA 2602
Qy 2630 GCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGG 2689
Db 2603 GGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACCCCAGT 2659
Qy 2690 ATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCA 2749
Db 2660 GTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGAACATA 2717
Qy 2750 GACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db 2718 CACTGTGAACTCAATCTCCAGGGCTCCAAAT 2748

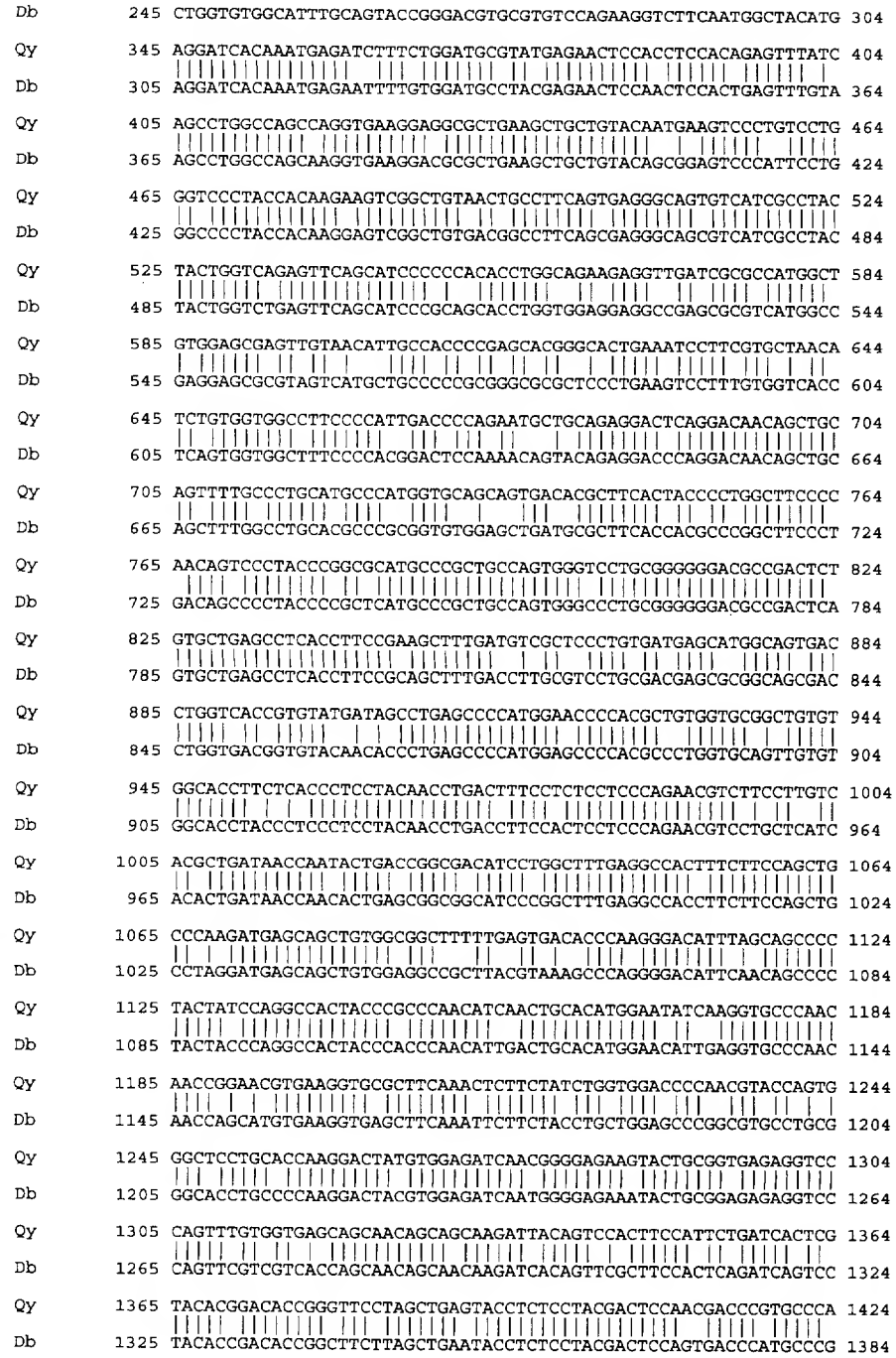
RESULT 3
US-09-776-191-1
; Sequence 1, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiunn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; OTHER INFORMATION: Nucleotide encoding MTSP1
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank AR081724
; PATENT DOCUMENT NUMBER: US Pat 5972616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-09-776-191-1

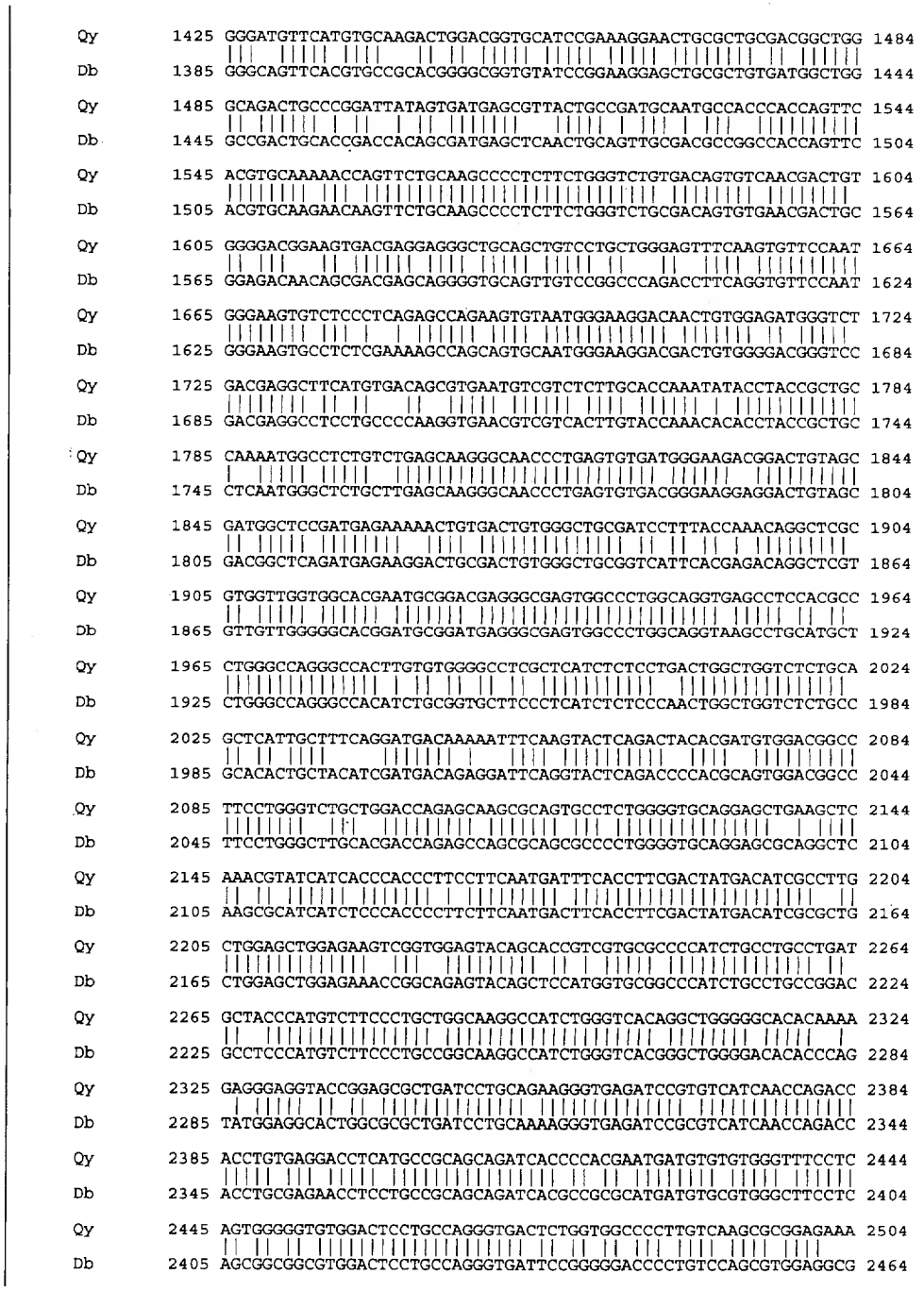
Query Match 60.6%; Score 1883.2; DB 11; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

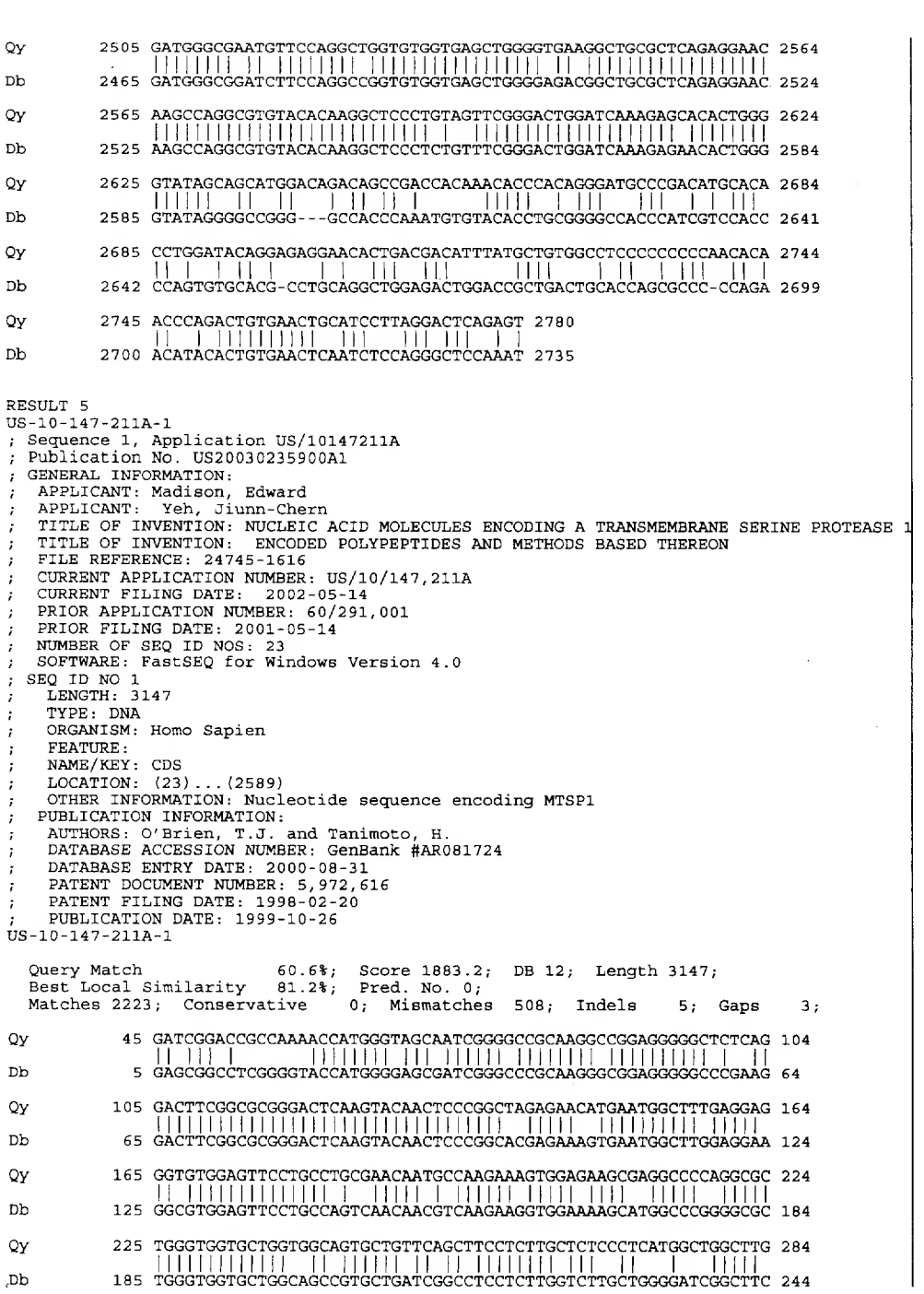
```
Qy        45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
             || ||| |       |||||||| ||| |||||| |||||||| |||||||||| |   ||
Db         5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy       105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
             ||||||||||||||||||||||||||||||||  |||||  |||||||||| |||||
Db        65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy       165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
             ||  |||||||||||||  |   |||||  | ||||||  |||||  ||||   |||||  |||||
Db       125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy       225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
             ||||||||||||   ||  ||||||  ||  ||  ||||||| |||   ||    |  |||||
Db       185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy       285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
             ||||||||||| ||  ||  ||  |||  |  |||||  ||  ||  ||  ||||||||||||  |   ||
Db       245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy       345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
             ||||||||||||||   |||  |||||||  ||  ||||||||||  ||||||  ||||||  |
Db       305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy       405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
```

```
Db       365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy       465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
             ||  ||||||||||| |||||||||  ||  |||||||  |||||||  ||||||||||||
Db       425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy       525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
             |||||||  |||||||||||||  |  |||||||  ||  ||||  ||  ||||  ||||||
Db       485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy       585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
             |  ||||||  ||  ||  |   ||||  ||  ||  ||  ||  |  |||||  |||||  |||  |  ||
Db       545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy       645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
             ||  ||||||||  |||||||   |||  |||  ||   |  ||||||||  ||||||||||||||||
Db       605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy       705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
             ||  ||||  ||||||  ||||   ||||  |  |||  ||||||||  ||  ||  ||||||||
Db       665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy       765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCT 824
             ||||  ||||||||  ||  ||||||||||||||||||||  ||||||||||||||||||||
Db       725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGGACGCCGACTCA 784

Qy       825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
             ||||||||||||||||||  |||||||  |  ||  ||||  ||  ||||  |||||  |||
Db       785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy       885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
             |||||  ||  |||||  |  |  ||||||||||||||||  |||||||  ||||||  |  |||||
Db       845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy       945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
             |||||||  |  ||||||||||||||||  ||||  ||||||||||||||||  |  ||  ||
Db       905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy      1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
             ||  ||||||||||  |||||  |||||  |||||  ||||||||||||||||  ||||||||||
Db       965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy      1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
             ||  |  ||||||||||||||  |||   ||  ||  |  ||||  ||||||||  |  |||||||
Db      1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy      1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
             |||||  ||||||||||||||  |||||||  ||||||||||||  ||  ||||||||||
Db      1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy      1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
             ||||  |  |||||||||  |||||||  |||  ||||  |||   |||  ||  ||  |  |
Db      1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy      1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
             |||  |||||  ||||||||||  ||||||||||  |||||||  |||||||  |||||||||
Db      1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy      1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
             ||||  ||  |  ||||||||||  ||||||  ||||  |  ||||||  ||  |||||  ||
Db      1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy      1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
             ||||  ||||||||  |||  |||||||  ||||||||||||||||||  ||||||  ||||||
Db      1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy      1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
             |||  ||||  |||  ||  ||  |||||  ||||  |||||  ||||||||  ||  ||||||
Db      1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy      1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
             ||  ||||||  |  ||  ||  |||||||  ||||  |  |||  |  |||  ||||||||||
```

```
Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 4
US-09-776-191-49
; Sequence 49, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
;  APPLICANT: Edwin L. Madison
;  APPLICANT:  Edgar O. Ong
;  APPLICANT:  Jiunn-Chern Yeh
;  APPLICANT:  Corvas International, Inc.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
;  TITLE OF INVENTION:  TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
;  TITLE OF INVENTION:  METHODS BASED THEREON
;  FILE REFERENCE: 24745-1607
;  CURRENT APPLICATION NUMBER: US/09/776,191
;  CURRENT FILING DATE:  2001-02-02
;  PRIOR APPLICATION NUMBER: 60/213,124
;  PRIOR FILING DATE: 2000-06-22
;  PRIOR APPLICATION NUMBER: 60/234,840
;  PRIOR FILING DATE: 2000-06-22
;  PRIOR APPLICATION NUMBER: 60/179,982
;  PRIOR FILING DATE: 2000-02-03
;  PRIOR APPLICATION NUMBER: 60/183,542
;  PRIOR FILING DATE: 2000-02-18
;  PRIOR APPLICATION NUMBER: 09/657,968
;  PRIOR FILING DATE: 2000-02-08
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1865)...(2590)
;   OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-09-776-191-49

  Query Match            60.6%;  Score 1883.2;  DB 11;  Length 3147;
  Best Local Similarity  81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
```

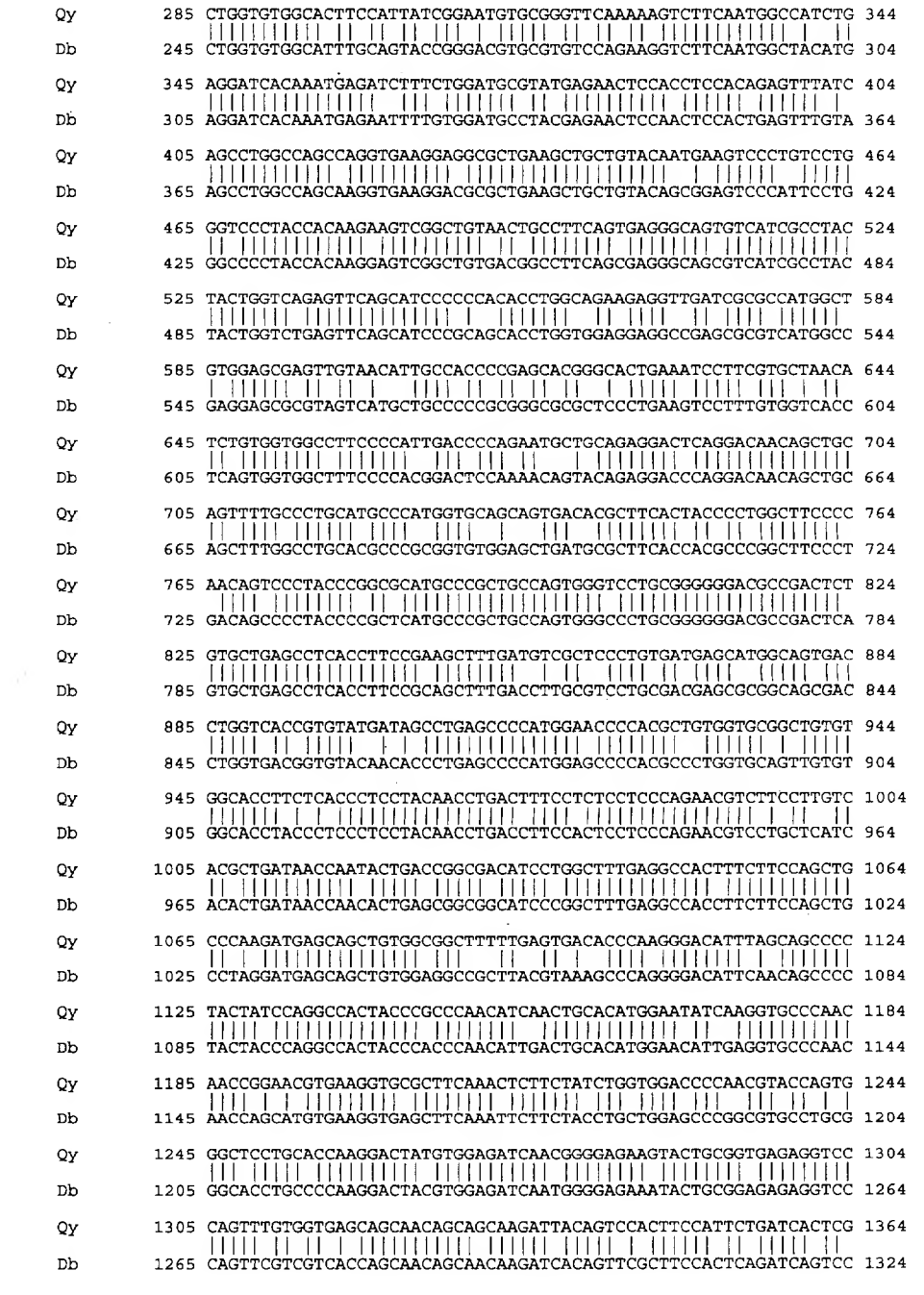






RESULT 5
US-10-147-211A-1
; Sequence 1, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edward
; APPLICANT: Yeh, Jiunn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/291,001
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSP1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-147-211A-1

Query Match 60.6%; Score 1883.2; DB 12; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

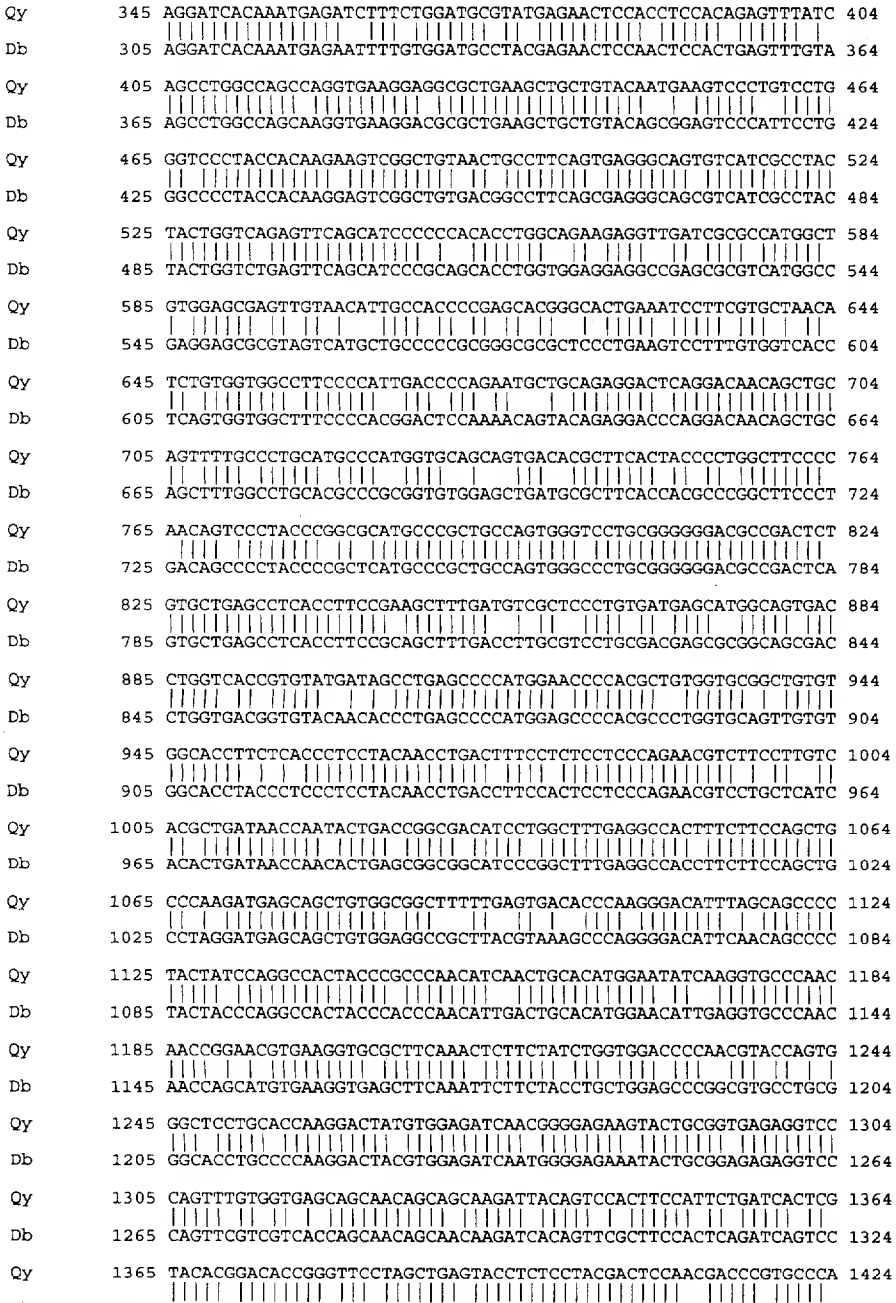


```
Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504

                 || || || |||||||||||||||||||| || || ||| |||| |||| ||||
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 6
US-10-147-211A-3
; Sequence 3, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edward
; APPLICANT: Yeh, Jiunn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1(
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/291,001
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-10-147-211A-3

Query Match 60.6%; Score 1883.2; DB 12; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

```
Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304
```


Qy 345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364
Qy 405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db 365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424
Qy 465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db 425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484
Qy 525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db 485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544
Qy 585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db 545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604
Qy 645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664
Qy 705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db 665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724
Qy 765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db 725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784
Qy 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844
Qy 885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db 845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904
Qy 945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db 905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964
Qy 1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db 965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024
Qy 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084
Qy 1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144
Qy 1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db 1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204
Qy 1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264
Qy 1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db 1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
Qy 1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db 1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
Qy 1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
Qy 1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db 1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504
Qy 1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564
Qy 1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db 1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624
Qy 1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744
Qy 1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db 1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804
Qy 1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864
Qy 1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db 1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924
Qy 1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984
Qy 2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044
Qy 2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104
Qy 2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164
Qy 2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224
Qy 2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284
Qy 2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db 2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344
Qy 2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db 2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404
Qy 2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db 2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464
Qy 2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

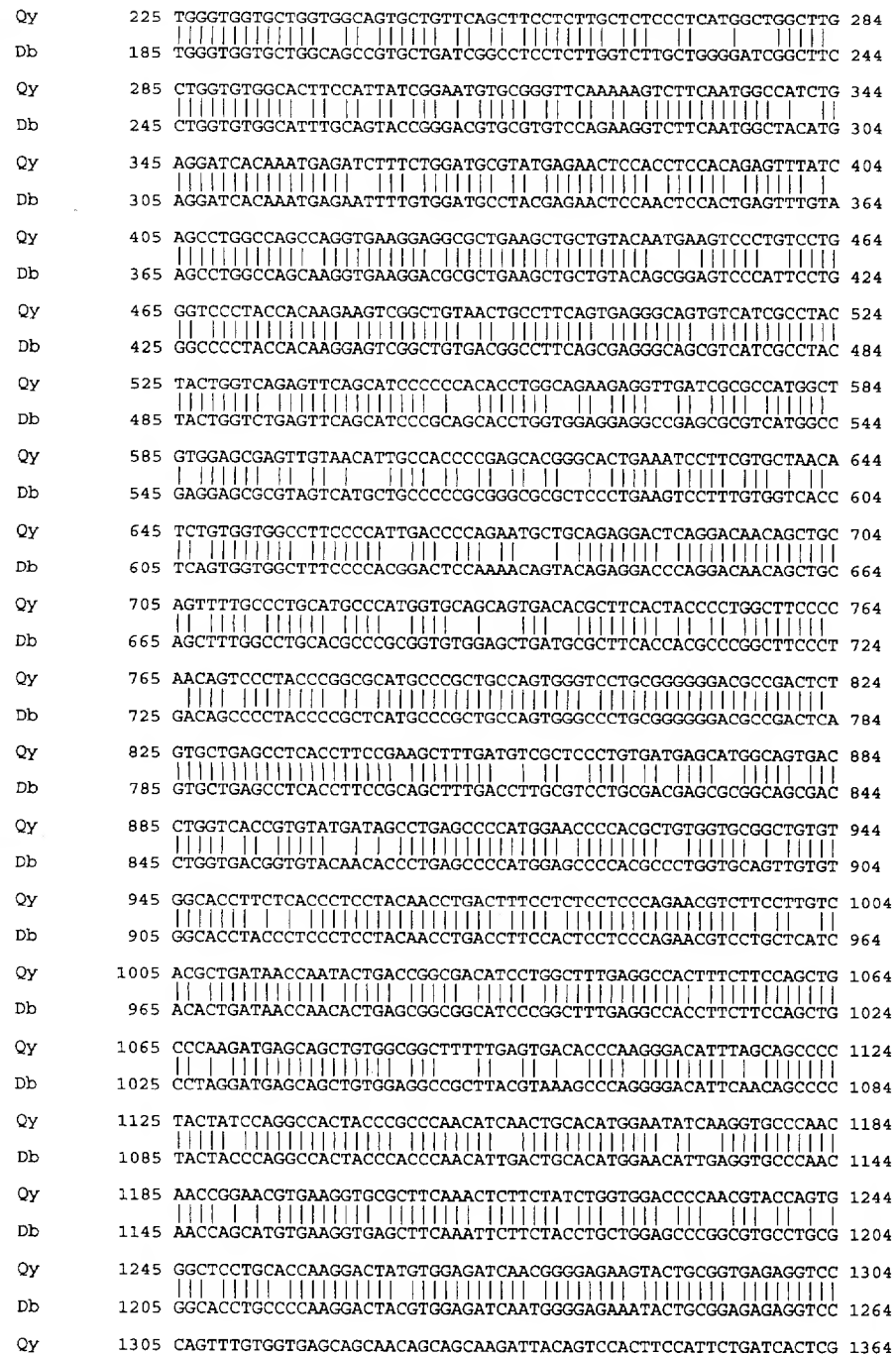
```
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 7
US-10-156-214A-1
; Sequence 1, Application US/10156214A
; Publication No. US20040001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Semple
; APPLICANT: George P. Vlasuk
; APPLICANT: Scott Jeffrey Kemp
; APPLICANT: Mallareddy Komandla
; APPLICANT: Daniel Vanna Siev
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Us
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 24745-1611
; CURRENT APPLICATION NUMBER: US/10/156,214A
; CURRENT FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSP1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: US Pat 5972616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-156-214A-1

Query Match 60.6%; Score 1883.2; DB 12; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

```
Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104

Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164

Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224

Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284

Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404

Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464

Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524

Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584

Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644

Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704

Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764

Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824

Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884

Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944

Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004

Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064

Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124

Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184

Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244

Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304

Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364

Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
```

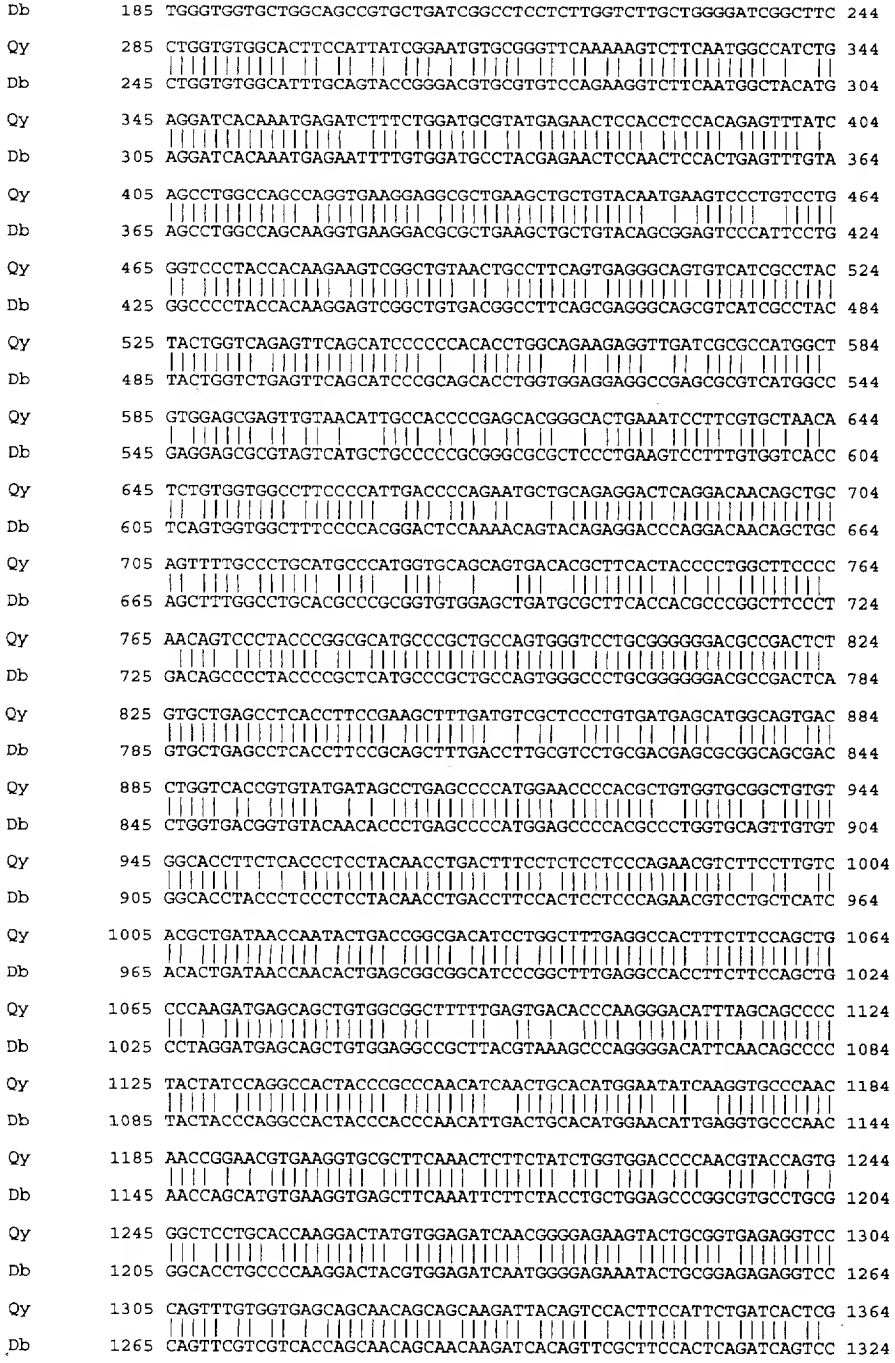
```
Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
             ||||| |||||||| ||| ||||||| ||||||||||||||||||||  ||||| |||||
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
             |||  ||||| |||    || || ||||| ||||| ||||| |||||||| || ||||||
Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
             || |||||| | ||  | || |||||||    |||||  | |||  | |||  ||||||||||
Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
             |||||||| ||| |||||||||||||||||||||||||||| ||||||||  ||||||||
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
             || |||  || ||||||  |||| ||||| ||||| ||    ||  |||| ||||||||||
Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
             |||||||| ||| |  | |||||| |||| |||||||||||| |||||||  || |||||
Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
             |||||||| || ||   ||  ||||| |||||| |||| |||||| | ||||||||||||
Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
             |  ||||| |||||  |||||||||||||||||||||||||||||| ||||||  ||||||||
Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
             || ||||| |||||||||  |||| |||||||||||||| || || || | |||||||||
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
             || ||||| |||||| ||||||| |||||||||||||||||||||| ||||| || ||
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
             |||||||||||||| | || || || |||||||||||  ||||||||||||||||||||
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
             || || ||||      |||||| |  |||| |||||||||||  ||||  ||||||||||
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
             ||||||||  ||| |||||||| |||||| ||| ||||||||||||||  | ||||
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
             || || |||||| |||||| | |||||||| |||||||||||||||||||||||||||  ||
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
             |||||||||||||  |||  |||||||| || | |||||  |||||||||||||||  ||
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
             ||  |||||||||||||| |||||||||||||||||||| |||||||  |||||  |
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
             | ||||| || || |||||||||||||| |||||||||||||| ||||||||||||||||||
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
             ||||| ||| ||||| |||||||||||||||| || || |||||||| ||||| ||||||
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
             || || || |||||||||||||||||||||||||| || || || ||| |||| |||| ||||
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
             |||||||| || |||||||| |||||||||||||||||| || ||||||||||||||||||
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
             |||||||||||||||||||||||||||||| | ||||||||||||||||||| ||||||||
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
             ||||||  ||  ||   | || || |     |||||  | |||   |||  | | |||
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
             || |  | || |    | |  |||  |||       ||||    | ||  | |||  || |
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
             ||  | ||||||||||  |||   ||| |||  | |
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 8
US-10-190-030B-1
; Sequence 1, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
;  APPLICANT: Madison, Edwin
;  APPLICANT:  Ong, Edgar
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
;  TITLE OF INVENTION:  TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
;  TITLE OF INVENTION:  METHODS BASED THEREON
;  FILE REFERENCE: 24745-1618
;  CURRENT APPLICATION NUMBER: US/10/190,030B
;  CURRENT FILING DATE:  2002-07-03
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (23)...(2589)
;   OTHER INFORMATION: Nucleotide sequence encoding MTSP1
;  PUBLICATION INFORMATION:
;   AUTHORS: O'Brien, T.J. and Tanimoto, H.
;   DATABASE ACCESSION NUMBER: GenBank #AR081724
;   DATABASE ENTRY DATE: 2000-08-31
;   PATENT DOCUMENT NUMBER: 5,972,616
;   PATENT FILING DATE: 1998-02-20
;   PUBLICATION DATE: 1999-10-26
US-10-190-030B-1

  Query Match           60.6%;  Score 1883.2;  DB 13;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
              || ||| |       |||||||| ||| |||||| |||||||| |||||||||  |  ||
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
              |||||||||||||||||||||||||||||||||||||||||||||  |||||  |||||||||
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
              || |||||||||||||  |||||  | ||||||  |||||  ||||  |||||  |||||
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184
```


Qy 225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db 185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244
Qy 285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304
Qy 345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364
Qy 405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db 365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424
Qy 465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db 425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484
Qy 525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db 485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544
Qy 585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db 545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604
Qy 645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664
Qy 705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db 665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724
Qy 765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db 725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784
Qy 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844
Qy 885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db 845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904
Qy 945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db 905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964
Qy 1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db 965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024
Qy 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084
Qy 1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144
Qy 1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db 1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204
Qy 1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264
Qy 1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db 1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
Qy 1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db 1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
Qy 1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
Qy 1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db 1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504
Qy 1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564
Qy 1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db 1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624
Qy 1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744
Qy 1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db 1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804
Qy 1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864
Qy 1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db 1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924
Qy 1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984
Qy 2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044
Qy 2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104
Qy 2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164
Qy 2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224
Qy 2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284
Qy 2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db 2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344
Qy 2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444

```
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAA 2504

Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 9
US-10-190-030B-3
; Sequence 3, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
;  APPLICANT: Madison, Edwin
;  APPLICANT:  Ong, Edgar
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
;  TITLE OF INVENTION:  TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
;  TITLE OF INVENTION:  METHODS BASED THEREON
;  FILE REFERENCE: 24745-1618
;  CURRENT APPLICATION NUMBER: US/10/190,030B
;  CURRENT FILING DATE:  2002-07-03
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1865)...(2590)
;   OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-10-190-030B-3

  Query Match             60.6%;  Score 1883.2;  DB 13;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104

Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164

Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224

Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284

Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404

Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464

Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524

Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584

Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644

Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704

Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764

Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824

Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884

Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944

Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004

Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064

Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124

Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184

Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244

Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304

Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364

Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
```

```
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
             |||  ||||| ||||   || || ||||| ||||| |||||| || ||||||
Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy       1485 GCAGACTGCCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
             || |||||| | ||  | || |||||||   ||||| | ||| | |||  ||||||||||
Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
             |||||||| ||| |||||||||||||||||||||||||||| |||||||| ||||||||
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
             || ||| || |||||| |||| ||||| ||||| ||   ||  |||| ||||||||||
Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
             |||||||| ||| |  | |||||| |||| |||||||||||| ||||||| || |||||
Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
             |||||||| || ||   || ||||| |||||| |||| |||||| | ||||||||||||
Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
             |  ||||| ||||| ||||||||||||||||||||||||||||| ||||||  ||||||||||
Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
             || ||||| |||||||| |||| ||||||||||||||| || || || | |||||||||
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
             || ||||| |||||| ||||||| |||||||||||||||||||||| ||||| || ||
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
             |||||||||||||| | || || || || |||||||||| ||||||||||||||||||
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
             || || |||| |||||||| |  |||| |||||||||  ||||  ||||||||||
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
             |||||||  |||  |||||||| ||||||| ||| |||||||||||||  | ||||
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
             || || |||||| ||||||| | |||||||| |||||||||||||||||||||||||  ||
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
             |||||||||||||  |||  |||||||| || | ||||| ||||||||||||||| ||
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
             ||  |||||||||||||| |||||||||||||||||||| |||||||| |||||  |
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
             | ||||| || ||||||||||||| |||||||||||||| |||||||||||||||||||
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
             ||||| ||| ||||| |||||||||||||||| || || |||||||| ||||| ||||||
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
             || || || |||||||||||||||||| || || || ||| |||| |||| ||||
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464
```

```
Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
             |||||||| || |||||||| |||||||||||||||| || ||||||||||||||||||
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
             |||||||||||||||||||||||||||| | |||||||||||||||||| ||||||||
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
             |||||| ||  ||   | || || |      |||||  | |||   |||  | | |||
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
             || |  | || |·   | |  |||  |||      ||||    | ||  | |||  || |
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
             ||  | ||||||||||  |||   ||| |||  | |
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 10
US-10-267-219-1
; Sequence 1, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
;  APPLICANT: Madison, Edwin
;  APPLICANT:  Yeh, Jiunn-Chern
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2
;  TITLE OF INVENTION:  ENCODED POLYPEPTIDES AND METHODS BASED THEREON
;  FILE REFERENCE: 24745-1621
;  CURRENT APPLICATION NUMBER: US/10/267,219
;  CURRENT FILING DATE:  2002-10-08
;  PRIOR APPLICATION NUMBER: 60/328,530
;  PRIOR FILING DATE: 09-OCT-2001
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (23)...(2589)
;   OTHER INFORMATION: Nucleotide sequence encoding MTSP1
;  PUBLICATION INFORMATION:
;   AUTHORS: O'Brien, T.J. and Tanimoto, H.
;   DATABASE ACCESSION NUMBER: GenBank #AR081724
;   DATABASE ENTRY DATE: 2000-08-31
;   PATENT DOCUMENT NUMBER: 5,972,616
;   PATENT FILING DATE: 1998-02-20
;   PUBLICATION DATE: 1999-10-26
US-10-267-219-1

  Query Match             60.6%;  Score 1883.2;  DB 13;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
              || ||| |       |||||||| ||| |||||| |||||||| |||||||||| |  ||
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
              |||||||||||||||||||||||||||||||| |||||  |||||||||| |||||
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
              || |||||||||||||  |||| | |||||| |||| ||||  |||||  |||||
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
              |||||||||||||  || |||||| || || |||||||| |||  ||   |  |||||
```


Db 185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244
Qy 285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db 245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304
Qy 345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364
Qy 405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db 365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424
Qy 465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db 425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484
Qy 525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db 485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544
Qy 585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db 545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604
Qy 645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664
Qy 705 AGTTTTGCCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db 665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724
Qy 765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db 725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784
Qy 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844
Qy 885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db 845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904
Qy 945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db 905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964
Qy 1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db 965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024
Qy 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084
Qy 1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144
Qy 1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db 1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204
Qy 1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264
Qy 1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db 1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
Qy 1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db 1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
Qy 1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
Qy 1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db 1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504
Qy 1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564
Qy 1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db 1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624
Qy 1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744
Qy 1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db 1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804
Qy 1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864
Qy 1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db 1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924
Qy 1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984
Qy 2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044
Qy 2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104
Qy 2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164
Qy 2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224
Qy 2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284
Qy 2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db 2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344
Qy 2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db 2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

```
Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAA 2504
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

```
RESULT 11
US-10-267-219-3
; Sequence 3, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
;  APPLICANT: Madison, Edwin
;  APPLICANT:  Yeh, Jiunn-Chern
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2
;  TITLE OF INVENTION:  ENCODED POLYPEPTIDES AND METHODS BASED THEREON
;  FILE REFERENCE: 24745-1621
;  CURRENT APPLICATION NUMBER: US/10/267,219
;  CURRENT FILING DATE:  2002-10-08
;  PRIOR APPLICATION NUMBER: 60/328,530
;  PRIOR FILING DATE: 09-OCT-2001
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1865)...(2590)
;   OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-10-267-219-3

  Query Match             60.6%;  Score 1883.2;  DB 13;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
```

```
Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
```

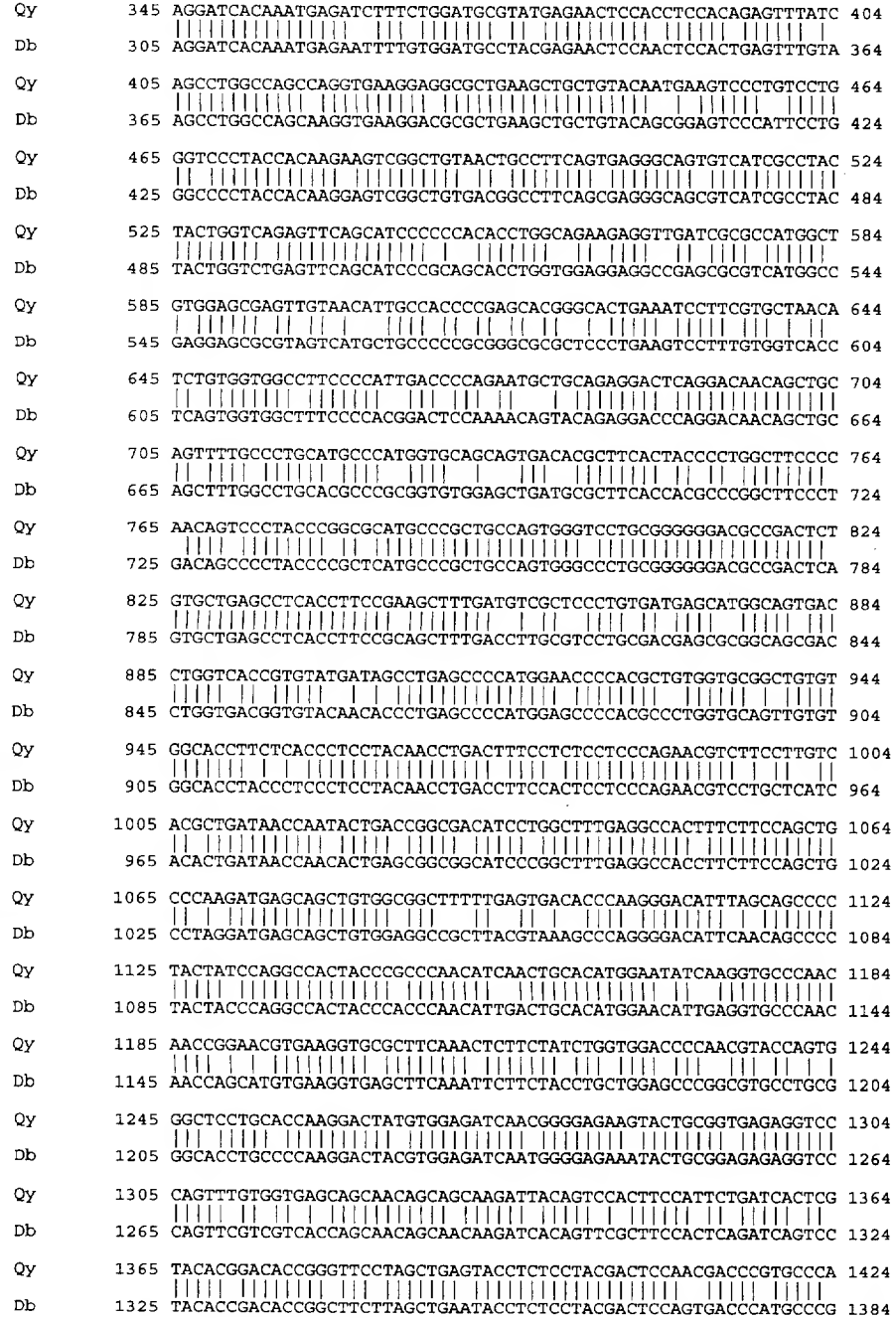
```
Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464
```

```
Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 12
US-10-302-840A-1
; Sequence 1, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCODI
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSP1
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
US-10-302-840A-1

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

```
Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304
```


Qy 345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db 305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364
Qy 405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db 365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424
Qy 465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db 425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484
Qy 525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db 485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544
Qy 585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db 545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604
Qy 645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db 605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664
Qy 705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db 665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724
Qy 765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db 725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784
Qy 825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db 785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844
Qy 885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db 845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904
Qy 945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db 905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964
Qy 1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db 965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024
Qy 1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db 1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084
Qy 1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db 1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144
Qy 1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db 1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204
Qy 1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db 1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264
Qy 1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db 1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324
Qy 1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db 1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384
Qy 1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
Db 1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
Qy 1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db 1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504
Qy 1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db 1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564
Qy 1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db 1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624
Qy 1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db 1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684
Qy 1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db 1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744
Qy 1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db 1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804
Qy 1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db 1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864
Qy 1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db 1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924
Qy 1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db 1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984
Qy 2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db 1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044
Qy 2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db 2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104
Qy 2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db 2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164
Qy 2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db 2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224
Qy 2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db 2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284
Qy 2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db 2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344
Qy 2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db 2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404
Qy 2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAA 2504
Db 2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464
Qy 2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

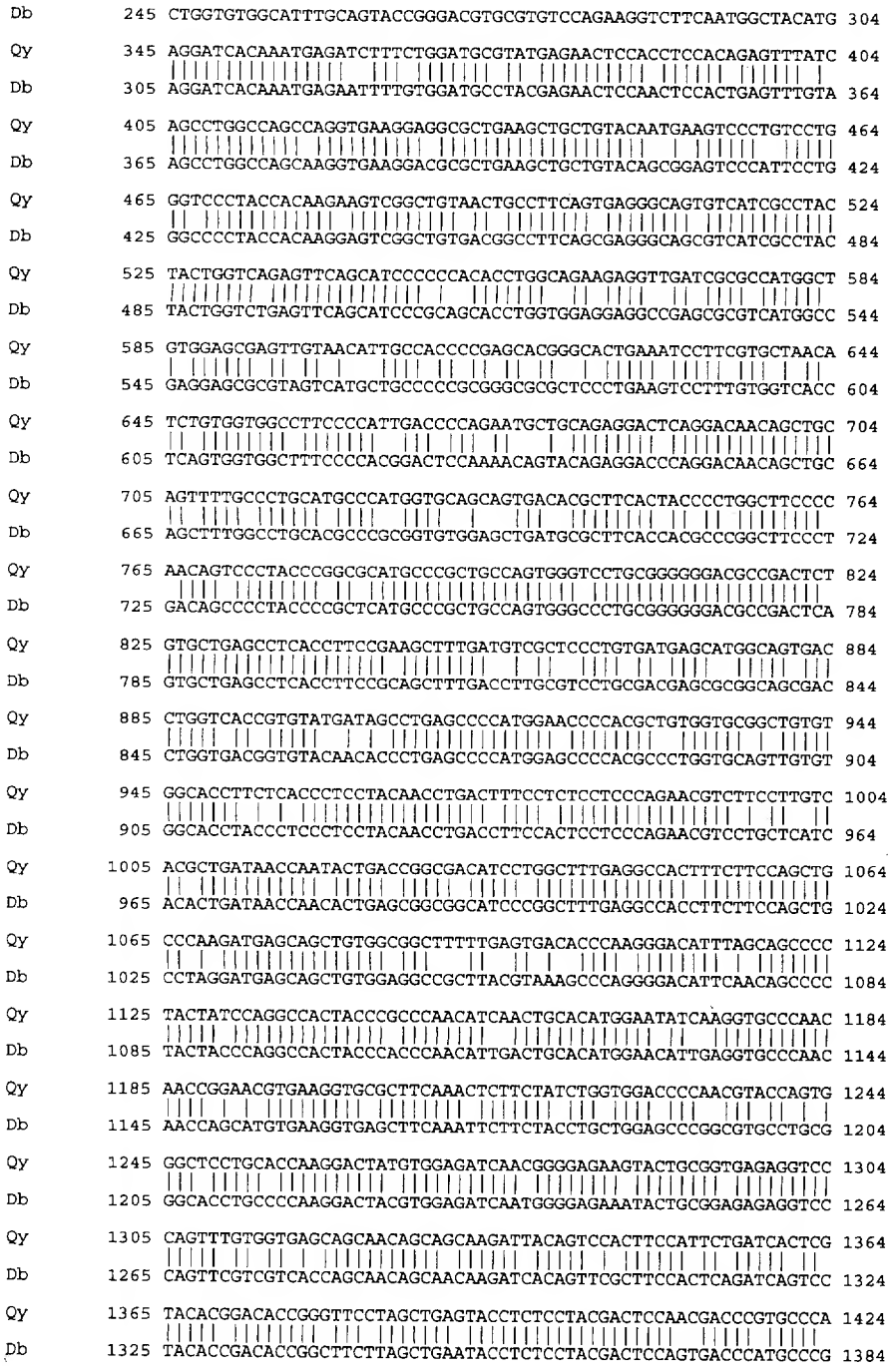
```
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 13
US-10-302-840A-3
; Sequence 3, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
;  APPLICANT: Madison, Edwin L.
;  APPLICANT:  Ong, Edgar O.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCOD
;  TITLE OF INVENTION:  POLYPEPTIDES AND METHODS BASED THEREON
;  FILE REFERENCE: 24745-1622
;  CURRENT APPLICATION NUMBER: US/10/302,840A
;  CURRENT FILING DATE:  2003-01-24
;  PRIOR APPLICATION NUMBER: 60/332,015
;  PRIOR FILING DATE: 2001-11-20
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1865)...(2590)
;   OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-10-302-840A-3

  Query Match           60.6%;  Score 1883.2;  DB 13;  Length 3147;
  Best Local Similarity  81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104

Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164

Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224

Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284

Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344

Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404

Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464

Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524

Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584

Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644

Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704

Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764

Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824

Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884

Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944

Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004

Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064

Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124

Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184

Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244

Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304

Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364

Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424

Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484

Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444
```

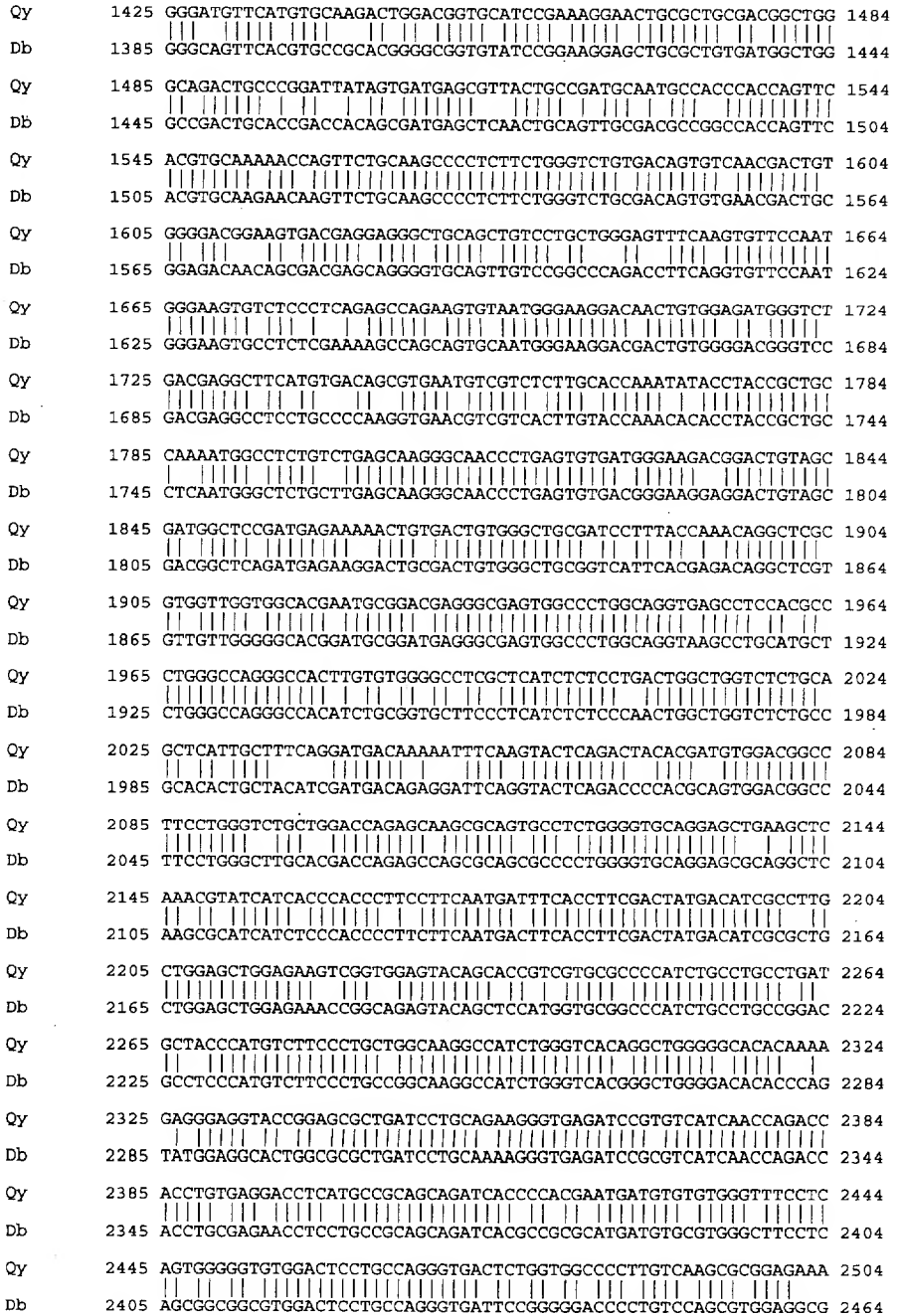
```
Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544
Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604
Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664
Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724
Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784
Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844
Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904
Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964
Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024
Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084
Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144
Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204
Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGAT 2264
Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324
Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384
Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444
Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504
Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

RESULT 14
US-10-112-221A-1
; Sequence 1, Application US/10112221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9,
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSP1
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: GenBank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-112-221A-1

Query Match 60.6%; Score 1883.2; DB 13; Length 3147;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 2223; Conservative 0; Mismatches 508; Indels 5; Gaps 3;

```
Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
```





```
Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564
Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624
Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684
Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACA 2744
Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780
Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735


RESULT 15
US-10-112-221A-3
; Sequence 3, Application US/10112221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
;  APPLICANT: Madison, Edwin
;  APPLICANT:  Ong, Edgar O.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9
;  TITLE OF INVENTION:  ENCODED POLYPEPTIDES AND METHODS BASED THEREON
;  FILE REFERENCE: 24745-1615
;  CURRENT APPLICATION NUMBER: US/10/112,221A
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: 60/279,228
;  PRIOR FILING DATE: 2001-03-27
;  PRIOR APPLICATION NUMBER: 60/291,501
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 3147
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1865)...(2590)
;   OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSP1
US-10-112-221A-3

  Query Match             60.6%;  Score 1883.2;  DB 13;  Length 3147;
  Best Local Similarity   81.2%;  Pred. No. 0;
  Matches 2223;  Conservative    0;  Mismatches  508;  Indels    5;  Gaps    3;

Qy         45 GATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAG 104
Db          5 GAGCGGCCTCGGGGTACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAG 64

Qy        105 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAG 164
Db         65 GACTTCGGCGCGGGACTCAAGTACAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAA 124

Qy        165 GGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGC 224
Db        125 GGCGTGGAGTTCCTGCCAGTCAACAACGTCAAGAAGGTGGAAAAGCATGGCCCGGGGCGC 184

Qy        225 TGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTG 284
Db        185 TGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTCTTGGTCTTGCTGGGGATCGGCTTC 244

Qy        285 CTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTG 344
Db        245 CTGGTGTGGCATTTGCAGTACCGGGACGTGCGTGTCCAGAAGGTCTTCAATGGCTACATG 304

Qy        345 AGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATC 404
Db        305 AGGATCACAAATGAGAATTTTGTGGATGCCTACGAGAACTCCAACTCCACTGAGTTTGTA 364

Qy        405 AGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTG 464
Db        365 AGCCTGGCCAGCAAGGTGAAGGACGCGCTGAAGCTGCTGTACAGCGGAGTCCCATTCCTG 424

Qy        465 GGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTAC 524
Db        425 GGCCCCTACCACAAGGAGTCGGCTGTGACGGCCTTCAGCGAGGGCAGCGTCATCGCCTAC 484

Qy        525 TACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCT 584
Db        485 TACTGGTCTGAGTTCAGCATCCCGCAGCACCTGGTGGAGGAGGCCGAGCGCGTCATGGCC 544

Qy        585 GTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACA 644
Db        545 GAGGAGCGCGTAGTCATGCTGCCCCCGCGGGCGCGCTCCCTGAAGTCCTTTGTGGTCACC 604

Qy        645 TCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGC 704
Db        605 TCAGTGGTGGCTTTCCCCACGGACTCCAAAACAGTACAGAGGACCCAGGACAACAGCTGC 664

Qy        705 AGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCC 764
Db        665 AGCTTTGGCCTGCACGCCCGCGGTGTGGAGCTGATGCGCTTCACCACGCCCGGCTTCCCT 724

Qy        765 AACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCT 824
Db        725 GACAGCCCCTACCCCGCTCATGCCCGCTGCCAGTGGGCCCTGCGGGGGGACGCCGACTCA 784

Qy        825 GTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGAC 884
Db        785 GTGCTGAGCCTCACCTTCCGCAGCTTTGACCTTGCGTCCTGCGACGAGCGCGGCAGCGAC 844

Qy        885 CTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGT 944
Db        845 CTGGTGACGGTGTACAACACCCTGAGCCCCATGGAGCCCCACGCCCTGGTGCAGTTGTGT 904

Qy        945 GGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTC 1004
Db        905 GGCACCTACCCTCCCTCCTACAACCTGACCTTCCACTCCTCCCAGAACGTCCTGCTCATC 964

Qy       1005 ACGCTGATAACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTG 1064
Db        965 ACACTGATAACCAACACTGAGCGGCGGCATCCCGGCTTTGAGGCCACCTTCTTCCAGCTG 1024

Qy       1065 CCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCC 1124
Db       1025 CCTAGGATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCC 1084

Qy       1125 TACTATCCAGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAAC 1184
Db       1085 TACTACCCAGGCCACTACCCACCCAACATTGACTGCACATGGAACATTGAGGTGCCCAAC 1144

Qy       1185 AACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTG 1244
Db       1145 AACCAGCATGTGAAGGTGAGCTTCAAATTCTTCTACCTGCTGGAGCCCGGCGTGCCTGCG 1204

Qy       1245 GGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCC 1304
Db       1205 GGCACCTGCCCCAAGGACTACGTGGAGATCAATGGGGAGAAATACTGCGGAGAGAGGTCC 1264

Qy       1305 CAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCG 1364
Db       1265 CAGTTCGTCGTCACCAGCAACAGCAACAAGATCACAGTTCGCTTCCACTCAGATCAGTCC 1324

Qy       1365 TACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCA 1424
Db       1325 TACACCGACACCGGCTTCTTAGCTGAATACCTCTCCTACGACTCCAGTGACCCATGCCCG 1384

Qy       1425 GGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGG 1484
```

```
Db       1385 GGGCAGTTCACGTGCCGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGG 1444

Qy       1485 GCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTC 1544

Db       1445 GCCGACTGCACCGACCACAGCGATGAGCTCAACTGCAGTTGCGACGCCGGCCACCAGTTC 1504

Qy       1545 ACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGT 1604

Db       1505 ACGTGCAAGAACAAGTTCTGCAAGCCCCTCTTCTGGGTCTGCGACAGTGTGAACGACTGC 1564

Qy       1605 GGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAAT 1664

Db       1565 GGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGGCCCAGACCTTCAGGTGTTCCAAT 1624

Qy       1665 GGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCT 1724

Db       1625 GGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGACGGGTCC 1684

Qy       1725 GACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGC 1784

Db       1685 GACGAGGCCTCCTGCCCCAAGGTGAACGTCGTCACTTGTACCAAACACACCTACCGCTGC 1744

Qy       1785 CAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGC 1844

Db       1745 CTCAATGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGGAAGGAGGACTGTAGC 1804

Qy       1845 GATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGC 1904

Db       1805 GACGGCTCAGATGAGAAGGACTGCGACTGTGGGCTGCGGTCATTCACGAGACAGGCTCGT 1864

Qy       1905 GTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCC 1964

Db       1865 GTTGTTGGGGGCACGGATGCGGATGAGGGCGAGTGGCCCTGGCAGGTAAGCCTGCATGCT 1924

Qy       1965 CTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCA 2024

Db       1925 CTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCCCAACTGGCTGGTCTCTGCC 1984

Qy       2025 GCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCC 2084

Db       1985 GCACACTGCTACATCGATGACAGAGGATTCAGGTACTCAGACCCCACGCAGTGGACGGCC 2044

Qy       2085 TTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTC 2144

Db       2045 TTCCTGGGCTTGCACGACCAGAGCCAGCGCAGCGCCCCTGGGGTGCAGGAGCGCAGGCTC 2104

Qy       2145 AAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTG 2204

Db       2105 AAGCGCATCATCTCCCACCCCTTCTTCAATGACTTCACCTTCGACTATGACATCGCGCTG 2164

Qy       2205 CTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCGCCCCCATCTGCCTGCCTGAT 2264

Db       2165 CTGGAGCTGGAGAAACCGGCAGAGTACAGCTCCATGGTGCGGCCCATCTGCCTGCCGGAC 2224

Qy       2265 GCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAA 2324

Db       2225 GCCTCCCATGTCTTCCCTGCCGGCAAGGCCATCTGGGTCACGGGCTGGGGACACACCCAG 2284

Qy       2325 GAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACC 2384

Db       2285 TATGGAGGCACTGGCGCGCTGATCCTGCAAAAGGGTGAGATCCGCGTCATCAACCAGACC 2344

Qy       2385 ACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTC 2444

Db       2345 ACCTGCGAGAACCTCCTGCCGCAGCAGATCACGCCGCGCATGATGTGCGTGGGCTTCCTC 2404

Qy       2445 AGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAA 2504

Db       2405 AGCGGCGGCGTGGACTCCTGCCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCG 2464

Qy       2505 GATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAAC 2564

Db       2465 GATGGGCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAAC 2524

Qy       2565 AAGCCAGGCGTGTACACAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGG 2624

Db       2525 AAGCCAGGCGTGTACACAAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGG 2584

Qy       2625 GTATAGCAGCATGGACAGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACA 2684

Db       2585 GTATAGGGGCCGGG---GCCACCCAAATGTGTACACCTGCGGGGCCACCCATCGTCCACC 2641

Qy       2685 CCTGGATACAGGAGAGGAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACA 2744

Db       2642 CCAGTGTGCACG-CCTGCAGGCTGGAGACTGGACCGCTGACTGCACCAGCGCCC-CCAGA 2699

Qy       2745 ACCCAGACTGTGAACTGCATCCTTAGGACTCAGAGT 2780

Db       2700 ACATACACTGTGAACTCAATCTCCAGGGCTCCAAAT 2735
```

Search completed: February 21, 2004, 10:23:01
Job time : 1070 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:06:37 ; Search time 6211 Seconds
(without alignments)
12154.195 Million cell updates/sec

Title: US-09-900-751-1
Perfect score: 3106
Sequence: 1 catggtagacggctgcccgg.....ttaaaaaaaaaaaaaaaaaa 3106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3102.8	99.9	3248	11	BC008514	BC008514 Mus muscu
2	3087.6	99.4	4017	11	AK052738	AK052738 Mus muscu
3	956.4	30.8	2689	11	AK085410	AK085410 Mus muscu
4	843.8	27.2	944	13	BQ885551	BQ885551 AGENCOURT
5	833.6	26.8	884	12	BI688367	BI688367 603316692
6	788.8	25.4	803	12	BI690615	BI690615 603314210
7	778.2	25.1	990	13	BQ951529	BQ951529 AGENCOURT
8	760.4	24.5	762	12	BG919617	BG919617 602823173
9	759.4	24.4	874	14	CB208872	CB208872 AGENCOURT
10	747.2	24.1	798	12	BI664081	BI664081 603289385
11	739.2	23.8	746	14	BY756896	BY756896 BY756896
12	739.2	23.8	846	12	BI416218	BI416218 602989878
13	735.8	23.7	1116	10	BF301347	BF301347 602029796
14	732	23.6	786	12	BI646380	BI646380 603276244
15	721	23.2	928	12	BG915045	BG915045 602814325
16	720.8	23.2	777	12	BI150180	BI150180 602848502
17	716	23.1	928	12	BI694441	BI694441 603348025
18	710.8	22.9	838	12	BI656113	BI656113 603284774
19	707.4	22.8	767	12	BI695744	BI695744 603345094
20	706.8	22.8	840	12	BI453178	BI453178 603173173
21	706.2	22.7	939	14	CA976536	CA976536 AGENCOURT
22	692.6	22.3	725	12	BI248406	BI248406 602967029
23	689.6	22.2	960	12	BI414245	BI414245 602986510
24	687.2	22.1	853	10	BF540669	BF540669 602052757
25	684.4	22.0	697	12	BI853355	BI853355 603379936
26	681.8	22.0	708	12	BI078327	BI078327 602872681
27	681.2	21.9	838	13	BU152438	BU152438 AGENCOURT
28	678.2	21.8	1035	10	BF119461	BF119461 601758388
29	671	21.6	671	12	BG861619	BG861619 602795392
30	665.4	21.4	836	13	BU517436	BU517436 AGENCOURT
31	658.6	21.2	728	12	BI112917	BI112917 602898206
32	650.4	20.9	927	10	BF578780	BF578780 602094810
33	647.4	20.8	649	12	BI697395	BI697395 603348134
34	643.8	20.7	698	14	BY756283	BY756283 BY756283
35	633.6	20.4	886	12	BI657843	BI657843 603284649
36	630.4	20.3	924	12	BI454477	BI454477 603171439
37	623.8	20.1	748	10	BF584458	BF584458 602098352
38	623.4	20.1	854	12	BI416184	BI416184 602989813
39	621.4	20.0	765	10	BF118535	BF118535 601754802
40	617.2	19.9	630	10	BF320582	BF320582 uz54g02.y
41	613	19.7	613	14	CA551053	CA551053 C0838D07-
42	612.4	19.7	868	13	BU179919	BU179919 AGENCOURT
43	610.8	19.7	698	12	BG870626	BG870626 602791606
44	610.4	19.7	802	12	BI149332	BI149332 602848871
45	608.8	19.6	790	12	BG968045	BG968045 602832425

ALIGNMENTS

RESULT 1
BC008514
LOCUS BC008514 3248 bp mRNA linear HTC 20-SEP-2002
DEFINITION Mus musculus, suppression of tumorigenicity 14 (colon carcinoma), clone IMAGE:3489914, mRNA.
ACCESSION BC008514
VERSION BC008514.1 GI:14250187
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3248)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 7 Row: h Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6677898
This clone has the following problem: no 5' EST match.

FEATURES Location/Qualifiers
source 1. .3248
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3489914"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCI_CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

BASE COUNT 734 a 918 c 928 g 668 t
ORIGIN

Query Match 99.9%; Score 3102.8; DB 11; Length 3248;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3104; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
Db        111 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 170

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
Db        171 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 230

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
Db        231 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 290

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
Db        291 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 350

Qy        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
Db        351 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 410

Qy        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
Db        411 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 470

Qy        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420
Db        471 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 530

Qy        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480
Db        531 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 590

Qy        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540
Db        591 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 650

Qy        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600
Db        651 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 710

Qy        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660
Db        711 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 770

Qy        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720
Db        771 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 830

Qy        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780
Db        831 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 890

Qy        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840
Db        891 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 950

Qy        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900
Db        951 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 1010

Qy        901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960
Db       1011 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 1070

Qy        961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020
Db       1071 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1130

Qy       1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080
Db       1131 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1190

Qy       1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140
Db       1191 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1250

Qy       1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
Db       1251 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1310

Qy       1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
Db       1311 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1370

Qy       1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
Db       1371 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1430

Qy       1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
Db       1431 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1490

Qy       1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
Db       1491 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1550

Qy       1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
Db       1551 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1610

Qy       1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
Db       1611 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1670

Qy       1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
Db       1671 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1730

Qy       1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
Db       1731 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1790

Qy       1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
Db       1791 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1850
```

```
Qy       1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1851 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1910

Qy       1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1911 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1970

Qy       1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1971 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 2030

Qy       1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2031 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 2090

Qy       1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2091 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2150

Qy       2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2151 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2210

Qy       2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2211 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2270

Qy       2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2271 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2330

Qy       2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2331 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2390

Qy       2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2391 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2450

Qy       2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2451 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2510

Qy       2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2511 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2570

Qy       2461 CCTGCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2571 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2630

Qy       2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2631 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2690

Qy       2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2691 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2750

Qy       2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2751 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2810

Qy       2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2811 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACT 2870

Qy       2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2871 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2930

Qy       2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2931 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2990

Qy       2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2991 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 3050

Qy       2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3051 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3110

Qy       3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3111 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3170

Qy       3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAAAAAAAAAAAAAA 3106
             ||||||||||||||||||||||||| ||||||||||||||||||
Db       3171 TTGAGAGTAAAACATTTTATTTCTTTAAAAAAAAAAAAAAAAAAA 3216
```

RESULT 2
AK052738
LOCUS AK052738 4017 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone:D630041B06 product:suppression of tumorigenicity 14 (colon carcinoma), full insert sequence.
ACCESSION AK052738
VERSION AK052738.1 GI:26342936
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,

Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.

TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851

REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 4017)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.

FEATURES Location/Qualifiers
source 1. .4017
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:D630041B06"
/db_xref="taxon:10090"
/clone="D630041B06"
/tissue_type="kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
CDS 174. .2741
/note="unnamed protein product; putative suppression of tumorigenicity 14 (colon carcinoma) (MGD|MGI:1338881, GB|NM_011176, evidence: BLASTN, 100%, match=720)"
/codon_start=1
/protein_id="BAC35125.1"
/db_xref="GI:26342937"
/translation="MGSNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPANNA
KKVEKRGPRRWVVLVAVLFSFLLLSLMAGLLVWHFHYRNVRVQKVFNGHLRITNEIFL
DAYENSTSTEFISLASQVKEALKLLYNEVPVLGPYHKKSAVTAFSEGSVIAYYWSEFS
IPPHLAEEVDRAMAVERVVTLPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFAL
HAHGAAVTRFTTPGFPNSPYPAHARCQWVLRGDADSVLSLTFRSFDVAPCDEHGSDLV
TVYDSLSPMEPHAVVRLCGTFSPSYNLTFLSSQNVFLVTLITNTDRRHPGFEATFFQL
PKMSSCGGFLSDTQGTFSSPYYPGHYPPNINCTWNIKVPNNRNVKVRFKLFYLVDPNV
PVGSCTKDYVEINGEKYCGERSQFVVSSNSSKITVHFHSDHSYTDTGFLAEYLSYDSN
DPCPGMFMCKTGRCIRKELRCDGWADCPDYSDERYCRCNATHQFTCKNQFCKPLFWVC
DSVNDCGDGSDEEGCSCPAGSFKCSNGKCLPQSQKCNGKDNCGDGSDEASCDSVNVVS
CTKYTYRCQNGLCLSKGNPECDGKTDCSDGSDEKNCDCGLRSFTKQARVVGGTNADEG
EWPWQVSLHALGQGHLCGASLISPDWLVSAAHCFQDDKNFKYSDYTMWTAFLGLLDQS
KRSASGVQELKLKRIITHPSFNDFTFDYDIALLELEKSVEYSTVVRPICLPDATHVFP
AGKAIWVTGWGHTKEGGTGALILQKGEIRVINQTTCEDLMPQQITPRMMCVGFLSGGV
DSCQGDSGGPLSSAEKDGRMFQAGVVSWGEGCAQRNKPGVYTRLPVVRDWIKEHTGV"

BASE COUNT 861 a 1098 c 1164 g 894 t
ORIGIN

Query Match 99.4%; Score 3087.6; DB 11; Length 4017;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3090; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        112 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 171

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 231

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 291

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        292 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 351

Qy        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 411

Qy        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        412 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 471

Qy        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGG 531

Qy        421 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGA 591

Qy        481 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 AGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCA 651

Qy        541 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 GCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAA 711

Qy        601 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        712 CATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCC 771

Qy        661 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        772 CCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG 831

Qy        721 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        832 CCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGG 891

Qy        781 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        892 CGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCT 951

Qy        841 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 TCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATG 1011
```

```
Qy      901 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1012 ATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCT 1071

Qy      961 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1072 CCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATA 1131

Qy     1021 CTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCT 1080
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     1132 CTGACCGGCGACATCCTGGCTTTGAGGCCACCTTCTTCCAGCTGCCCAAGATGAGCAGCT 1191

Qy     1081 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1192 GTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACT 1251

Qy     1141 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1252 ACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGG 1311

Qy     1201 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1312 TGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGG 1371

Qy     1261 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1372 ACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCA 1431

Qy     1321 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1432 GCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGT 1491

Qy     1381 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1492 TCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCA 1551

Qy     1441 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1552 AGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATT 1611

Qy     1501 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1612 ATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGT 1671

Qy     1561 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1672 TCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACG 1731

Qy     1621 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1732 AGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTC 1791

Qy     1681 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1792 AGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTG 1851

Qy     1741 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1852 ACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTC 1911

Qy     1801 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1912 TGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGA 1971

Qy     1861 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1972 AAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGA 2031

Qy     1921 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2032 ATGCGGACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACT 2091

Qy     1981 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2092 TGTGTGGGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGG 2151

Qy     2041 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2152 ATGACAAAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGG 2211

Qy     2101 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2212 ACCAGAGCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCC 2271

Qy     2161 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2272 ACCCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGT 2331

Qy     2221 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2332 CGGTGGAGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCC 2391

Qy     2281 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2392 CTGCTGGCAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAG 2451

Qy     2341 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2452 CGCTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCA 2511

Qy     2401 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2512 TGCCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACT 2571

Qy     2461 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2572 CCTGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCC 2631

Qy     2521 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2632 AGGCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACA 2691

Qy     2581 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2692 CAAGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGAC 2751

Qy     2641 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2752 AGACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAG 2811

Qy     2701 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCAGACTGTGAACT 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2812 GAACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCAACACAACCCAGACTGTGAACT 2871

Qy     2761 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2872 GCATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAAC 2931

Qy     2821 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2932 TTGCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCC 2991

Qy     2881 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 2940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2992 CTGAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCC 3051

Qy     2941 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3052 GGGAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGA 3111

Qy     3001 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3112 ACCCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTAT 3171

Qy     3061 TTGAGAGTAAAACATTTTATTTCTTTTTAAAAAA 3094
```

```
                     ||||||||||||||||||||||||||||  ||
Db       3172 TTGAGAGTAAAACATTTTATTTCTTTTTAGGTAA 3205

RESULT 3
AK085410
LOCUS       AK085410               2689 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched
            library, clone:D630022P10 product:suppression of tumorigenicity 14
            (colon carcinoma), full insert sequence.
ACCESSION   AK085410
VERSION     AK085410.1  GI:26102691
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2689)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2689
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D630022P10"
                     /db_xref="taxon:10090"
                     /clone="D630022P10"
                     /tissue_type="kidney"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="0 day neonate"
     misc_feature    1. .2689
                     /note="suppression of tumorigenicity 14 (colon carcinoma)
                     (MGD|MGI:1338881, GB|NM_011176, evidence: BLASTN, 100%,
                     match=720)"
BASE COUNT      590 a    747 c    766 g    586 t
ORIGIN

  Query Match           30.8%;  Score 956.4;  DB 11;  Length 2689;
  Best Local Similarity   76.7%;  Pred. No. 1.7e-177;
  Matches 1415;  Conservative    0;  Mismatches    1;  Indels  428;  Gaps    3;

Qy          1 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 CATGGTAGACGGCTGCCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAA 169

Qy         61 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 CCATGGGTAGCAATCGGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGAC 229

Qy        121 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 TCAAGTACAACTCCCGGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGC 289

Qy        181 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 CTGCGAACAATGCCAAGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGG 349

Qy        241 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        350 CAGTGCTGTTCAGCTTCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCC 409

Qy        301 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 360
```

```
Db        410 ATTATCGGAATGTGCGGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGA 469
Qy        361 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAG-CCTGGCCAGCCAG 419
Db        470 TCTTTCTGGATGCGTATGAGAACTCCACCTCCACAGAGTTTATCAGCCCTGGCCAGCCAG 529
Qy        420 GTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAG 479
Db        530 GTGAAGGAGGCGCTGAAGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAG 589
Qy        480 AAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTC 539
Db        590 AAGTCGGCTGTAACTGCCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTC 649
Qy        540 AGCATCCCCCCACACCTGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTA 599
Db        650 AGCATCCCCCCACACCTGGCAAAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTA 709
Qy        600 ACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTC 659
Db        710 ACATTGCCACCCCGAGCACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTC 769
Qy        660 ------------------------------------------------------------ 659
Db        770 CGTGAGTGCGGGGCAAACCAGGGTGGGAATGTGTGGGCCAGCCTGTAAAGGGGCGGGGCT 829
Qy        660 ------------------------------------------------------------ 659
Db        830 GGGAGAGCCACACTCACAGGCCCAGGTGTCCTGGAGGACTCTGCTTGCTGAGTGGCACCT 889
Qy        660 ------------------------------------------------------------ 659
Db        890 CCTTTTGGGTAATAGGGAGGCAAGAGTAAAGAAATGATTGTCTGGGTAAAATGTTTGCCT 949
Qy        660 ------------CCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACA------------ 696
Db        950 TCATGTTCTCCACCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAGTAAGTATCAT 1009
Qy        697 ------------------------------------------------------------ 696
Db       1010 GGTGACCTCTGGTGGAGAAGAAAAGTGTCTTGCCCCTTAAGCCATGGGGTCCCTAAAAGT 1069
Qy        697 ------------------------------------------------------------ 696
Db       1070 CAGATGGTACTGATGGACAAGCATCTCTTCTTTTCCCCTTCCGGTCCAAAGCTCCTTTCC 1129
Qy        697 ------------------------------------------------------------ 696
Db       1130 TTGTGTCTACCTGAGTTCGCAGTGATCTGAGACGTATTGGGGGTAGAGCTAGGCGGGGGG 1189
Qy        697 ----------------------------------------------ACAGCTGCAGTTTTGC 712
Db       1190 CGGGGGCCTGCAGGGTTCACGCCTTCCTTCCATCTCTCCCACAGACAGCTGCAGTTTTGC 1249
Qy        713 CCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCC 772
Db       1250 CCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCC 1309
Qy        773 CTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAG 832
Db       1310 CTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAG 1369
Qy        833 CCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCAC 892
Db       1370 CCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCAC 1429
Qy        893 CGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTT 952
Db       1430 CGTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTT 1489
Qy        953 CTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGAT 1012
Db       1490 CTCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGAT 1549
Qy        1013 AACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGAT 1072
Db        1550 AACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGAT 1609
Qy        1073 GAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCC 1132
Db        1610 GAGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCC 1669
Qy        1133 AGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAA 1192
Db        1670 AGGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAA 1729
Qy        1193 CGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTG 1252
Db        1730 CGTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTG 1789
Qy        1253 CACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGT 1312
Db        1790 CACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGT 1849
Qy        1313 GGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGA 1372
Db        1850 GGTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGA 1909
Qy        1373 CACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACC 1416
Db        1910 CACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACC 1953
```

RESULT 4
BQ885551

```
LOCUS       BQ885551                944 bp    mRNA    linear   EST 16-AUG-2002
DEFINITION  AGENCOURT_8748027 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335237
            5', mRNA sequence.
ACCESSION   BQ885551
VERSION     BQ885551.1  GI:22277569
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 944)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13797  row: f  column: 06
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1. .944
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6335237"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_130"
                     /note="Organ: otocysts; Vector: pCMV-SPORT6.1.ccdb;
                     Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
                     Primer: Oligo dT.  Average insert size 1.95  kb.
                     Constructed by ResGen, Invitrogen Corp. Note: this is a
                     NIH_MGC Library."
BASE COUNT      184 a    288 c    272 g    200 t
```

ORIGIN

Query Match 27.2%; Score 843.8; DB 13; Length 944;
Best Local Similarity 96.3%; Pred. No. 2.7e-155;
Matches 885; Conservative 0; Mismatches 32; Indels 2; Gaps 2;

```
Qy         30 GCGTCTGAGACCGGCGATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCC 89
Db          1 GCGTCTGAGACCGGCGATCGGACCGCCAAAACCATGGGTAGCAATCGGGGCCGCAAGGCC 60

Qy         90 GGAGGGGGCTCTCAGGACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATG 149
Db         61 GGAGGGGGCTCTCAGGACTTCGGCGCGGGACTCAAGTACAACTCCCGGCTAGAGAACATG 120

Qy        150 AATGGCTTTGAGGAGGGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAG 209
Db        121 AATGGCTTTGAGGAGGGTGTGGAGTTCCTGCCTGCGAACAATGCCAAGAAAGTGGAGAAG 180

Qy        210 CGAGGCCCCAGGCGCTGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCC 269
Db        181 CGAGGCCCCAGGCGCTGGGTGGTGCTGGTGGCAGTGCTGTTCAGCTTCCTCTTGCTCTCC 240

Qy        270 CTCATGGCTGGCTTGCTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTC 329
Db        241 CTCATGGCTGGCTTGCTGGTGTGGCACTTCCATTATCGGAATGTGCGGGTTCAAAAAGTC 300

Qy        330 TTCAATGGCCATCTGAGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACC 389
Db        301 TTCAATGGCCATCTGAGGATCACAAATGAGATCTTTCTGGATGCGTATGAGAACTCCACC 360

Qy        390 TCCACAGAGTTTATCAGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAAT 449
Db        361 TCCACAGAGTTTATCAGCCTGGCCAGCCAGGTGAAGGAGGCGCTGAAGCTGCTGTACAAT 420

Qy        450 GAAGTCCCTGTCCTGGGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGC 509
Db        421 GAAGTCCCTGTCCTGGGTCCCTACCACAAGAAGTCGGCTGTAACTGCCTTCAGTGAGGGC 480

Qy        510 AGTGTCATCGCCTACTACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTT 569
Db        481 AGTGTCATCGCCTACTACTGGTCAGAGTTCAGCATCCCCCCACACCTGGCAGAAGAGGTT 540

Qy        570 GATCGCGCCATGGCTGTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAA 629
Db        541 GATCGCGCCATGGCTGTGGAGCGAGTTGTAACATTGCCACCCCGAGCACGGGCACTGAAA 600

Qy        630 TCCTTCGTGCTAACATCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACT 689
Db        601 TCCTTCGTGCTAACATCTGTGGTGGCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACT 660

Qy        690 CAGGACAACAGCTGCAGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACT 749
Db        661 CAGGACAACAGCTGCAGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACACGCTTCACT 720

Qy        750 ACCCCTGGCTTCCCCAACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGG 809
Db        721 ACCCCTGGCCTTCCCAACAGTCCCTACCCCGCGCATGCCCGCTGCCAGTGGGTCCTGCGG 780

Qy        810 GGGGACGCCGACTCTGTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGAT 869
Db        781 GGGGACGCCGACTCTGTGCTGAGCCTCACCTTCCGAACCTTTGATGTCCCTCCCTGTGAT 840

Qy        870 GAGCAT-GGCAGTGACCTGGTCACCGT-GTATGATAGCCTGAGCCCCATGGAACCCCACG 927
Db        841 GACCATGGGCATTGCCCTGGCCCCCGTGGTATGATAACCCTGAGCCCCATGGGACCCCCA 900

Qy        928 CTGTGGTGCGGCTGTGTGG 946
Db        901 CCCTGGTGGTGCGGGCTTG 919
```

RESULT 5
BI688367

```
LOCUS       BI688367                884 bp    mRNA    linear   EST 18-SEP-2001
DEFINITION  603316692F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5356487 5',
            mRNA sequence.
ACCESSION   BI688367
VERSION     BI688367.1  GI:15650996
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11906  row: h  column: 24
            High quality sequence stop: 831.
FEATURES             Location/Qualifiers
     source          1. .884
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5356487"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      197 a    232 c    268 g    187 t
ORIGIN
```

Query Match 26.8%; Score 833.6; DB 12; Length 884;
Best Local Similarity 98.6%; Pred. No. 2.7e-153;
Matches 872; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

```
Qy       1165 GGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGG 1224
Db          1 GGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATCTGG 60

Qy       1225 TGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGA 1284
Db         61 TGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGGAGA 120

Qy       1285 AGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCC 1344
Db         121 AGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAGTCC 180

Qy       1345 ACTTCCATTCTGATCACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACG 1404
Db         181 ACTTCCATTCTGATCACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCTACG 240

Qy       1405 ACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGG 1464
Db         241 ACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGG 300

Qy       1465 AACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGAT 1524
Db         301 AACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGAT 360

Qy       1525 GCAATGCCACCCACCAGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCT 1584
```

```
Db        361 GCAATGCCACCCACCAGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCT 420

Qy       1585 GTGACAGTGTCAACGACTGTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTG 1644

Db        421 GTGACAGTGTCAACGACTGTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTG 480

Qy       1645 GGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGG 1704

Db        481 GGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGG 540

Qy       1705 ACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCA 1764

Db        541 ACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCA 600

Qy       1765 CCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTGAG-CAAGGGCAACCCTGAGTGT 1823

Db        601 CCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTGAGCCAAGGGCAACCCTGAGTGT 660

Qy       1824 GATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGA 1883

Db        661 GATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGA 720

Qy       1884 TCCTTTACCAAACAGGCTCGCG-TGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCC 1942

Db        721 TCCTTTACCAGACAGGCTCGCGTTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCC 780

Qy       1943 CTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGC-TCATCT 2001

Db        781 CTGGCAGGTGAGCCTCCAAGCCCTGGGGCCAGGCCACTTGTGTGGGGGCTCGCTTCATCT 840

Qy       2002 CTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGAC 2045

Db        841 CTCCTGACTGGCTGGTCTCTGCAGCTCATCGCTTCCAGGATGCC 884
```

RESULT 6
BI690615

```
LOCUS       BI690615                803 bp    mRNA    linear   EST 18-SEP-2001
DEFINITION  603314210F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5354067 5',
            mRNA sequence.
ACCESSION   BI690615
VERSION     BI690615.1  GI:15653244
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11900  row: d  column: 04
            High quality sequence stop: 755.
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5354067"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      170 a    259 c    200 g    174 t
ORIGIN
```

Query Match 25.4%; Score 788.8; DB 12; Length 803;
Best Local Similarity 99.6%; Pred. No. 1.7e-144;
Matches 801; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy        654 GCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCC 713

Db          1 GCCTTCCCCATTGACCCCAGAATGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCC 60

Qy        714 CTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCC 773

Db         61 CTGCATGCCCATGGTGCAGCAGTGACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCC 120

Qy        774 TACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGC 833

Db        121 TACCCGGCGCATGCCCGCTGCCAGTGGGTCCTGCGGGGGGACGCCGACTCTGTGCTGAGC 180

Qy        834 CTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACC 893

Db        181 CTCACCTTCCGAAGCTTTGATGTCGCTCCCTGTGATGAGCATGGCAGTGACCTGGTCACC 240

Qy        894 GTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTC 953

Db        241 GTGTATGATAGCCTGAGCCCCATGGAACCCCACGCTGTGGTGCGGCTGTGTGGCACCTTC 300

Qy        954 TCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATA 1013

Db        301 TCACCCTCCTACAACCTGACTTTCCTCTCCTCCCAGAACGTCTTCCTTGTCACGCTGATA 360

Qy       1014 ACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATG 1073

Db        361 ACCAATACTGACCGGCGACATCCTGGCTTTGAGGCCACTTTCTTCCAGCTGCCCAAGATG 420

Qy       1074 AGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCA 1133

Db        421 AGCAGCTGTGGCGGCTTTTTGAGTGACACCCAAGGGACATTTAGCAGCCCCTACTATCCA 480

Qy       1134 GGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAAC 1193

Db        481 GGCCACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAAC 540

Qy       1194 GTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGC 1253

Db        541 GTGAAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGC 600

Qy       1254 ACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTG 1313

Db        601 ACCAAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTG 660

Qy       1314 GTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGAC 1373

Db        661 GTGAGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGAC 720

Qy       1374 ACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTC 1433

Db        721 ACCGGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATTGTC 780

Qy       1434 ATGTGCAAGACTGGACGGTGCATC 1457

Db        781 ATGTGCAAGA-TGGACGGTGCATC 803
```

RESULT 7
BQ951529

```
LOCUS       BQ951529                990 bp    mRNA    linear   EST 21-AUG-2002
```

```
DEFINITION  AGENCOURT_8952916 NCI_CGAP_Co24 Mus musculus cDNA clone
            IMAGE:6475074 5', mRNA sequence.
ACCESSION   BQ951529
VERSION     BQ951529.1  GI:22367007
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 990)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14013  row: d  column: 19
            High quality sequence stop: 635.
FEATURES             Location/Qualifiers
     source          1. .990
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6475074"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Co24"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      229 a    272 c    286 g    201 t      2 others
ORIGIN

  Query Match           25.1%;  Score 778.2;  DB 13;  Length 990;
  Best Local Similarity  92.9%;  Pred. No. 1.9e-142;
  Matches  924;  Conservative    0;  Mismatches   60;  Indels   11;  Gaps   10;

Qy        981 TCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATACTGACCGGCGACATCCTGGC 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATACTGACCGGCGACATCCTGGC 60

Qy       1041 TTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGAC 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGAC 120

Qy       1101 ACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACTACCCGCCCAACATCAACTGC 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACTACCCGCCCAACATCAACTGC 180

Qy       1161 ACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTAT 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTAT 240

Qy       1221 CTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGG 1280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGG 300

Qy       1281 GAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACA 1340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACA 360

Qy       1341 GTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCC 1400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCCACTTCCATTCTGATCACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCC 420

Qy       1401 TACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAGACTGGACGGTGCATCCGA 1460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAGACTGGACGGTGCATCCGA 480

Qy       1461 AAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGC 1520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGC 540

Qy       1521 CGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCT-CTTCTG 1579
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        541 CGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCTCCTTCTG 600

Qy       1580 GGTCTGTGACAGTGTCAACGACT-GTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTC 1638
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        601 GGTCTGTGACAGTGTCAACGACTGGTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTC 660

Qy       1639 CTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGT-GTAAT 1697
              |||||||||||||||||||| ||||||||||||||||||||||||||||||| |||||
Db        661 CTGCTGGGAGTTTCAAGTGTCCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGGTAAT 720

Qy       1698 GGGAAGGACAACTGT--GGAGATGGGTCTGACGAGGCTTCATGTG-ACAGCGTGAATGTC 1754
              |||||||| |||| ||||||||||||||||||||||||||||| ||||||||||||||
Db        721 GGGAAGGACNCCTGTTGGGAGATGGGTCTGACGAGGCTTCATGTGAACAGCGTGAATGTC 780

Qy       1755 GTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTGAGCAAGGGCAAC 1814
              |||||||||||| |||||||||||||||| |||||||| |||||||||||| |||||| |
Db        781 GTCTCTTGCACC-AATATACCTACCGCTGGCAAAATGG-CTCTGTCTGAGC-AGGGCACC 837

Qy       1815 CCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGT 1874
              | ||| || |||||||||||||||||| |||||| |||||||| |||||| | || |||
Db        838 CTTGATTGGGATGGGAAGACGGACTGTACCGATGG-TCCGATGAAAAAAAC-GGGAATGT 895

Qy       1875 GGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGC 1934
              |||||| || |||| | |||       | |       | | |       | || ||||
Db        896 GGGCTGGGACCCTTACCAAAAGGCTTCCCGGGTTGGTGGGCAAGCNCCCCCAACAAGGGG 955

Qy       1935 GAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGG 1969
              || ||||| |   | | | |  | || ||  ||
Db        956 CAGGGGCCCCGTGGAGGGGGACCTCCCGGCCGTGG 990


RESULT 8
BG919617
LOCUS       BG919617                762 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  602823173F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4951962 5',
            mRNA sequence.
ACCESSION   BG919617
VERSION     BG919617.1  GI:14300093
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 762)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10909  row: a  column: 19
            High quality sequence stop: 727.
FEATURES             Location/Qualifiers
     source          1. .762
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
```

```
                     /clone="IMAGE:4951962"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      164 a    243 c    190 g    165 t
ORIGIN

  Query Match             24.5%;  Score 760.4;  DB 12;  Length 762;
  Best Local Similarity   99.9%;  Pred. No. 6.4e-139;
  Matches  761;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        682 AGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACAC 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATGCCCATGGTGCAGCAGTGACAC 60

Qy        742 GCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGG 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGGCGCATGCCCGCTGCCAGTGGG 120

Qy        802 TCCTGCGGGGGGGACGCCGACTCTGTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTC 861
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCTGCGGGGGGACGCCGACTCTGTGCTGAGCCTCACCTTCCGAAGCTTTGATGTCGCTC 180

Qy        862 CCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAAC 921
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTGTGATGAGCATGGCAGTGACCTGGTCACCGTGTATGATAGCCTGAGCCCCATGGAAC 240

Qy        922 CCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCT 981
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCCACGCTGTGGTGCGGCTGTGTGGCACCTTCTCACCCTCCTACAACCTGACTTTCCTCT 300

Qy        982 CCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATACTGACCGGCGACATCCTGGCT 1041
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTCCCAGAACGTCTTCCTTGTCACGCTGATAACCAATACTGACCGGCGACATCCTGGCT 360

Qy       1042 TTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACA 1101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGAGGCCACTTTCTTCCAGCTGCCCAAGATGAGCAGCTGTGGCGGCTTTTTGAGTGACA 420

Qy       1102 CCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACTACCCGCCCAACATCAACTGCA 1161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCCAAGGGACATTTAGCAGCCCCTACTATCCAGGCCACTACCCGCCCAACATCAACTGCA 480

Qy       1162 CATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATC 1221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATGGAATATCAAGGTGCCCAACAACCGGAACGTGAAGGTGCGCTTCAAACTCTTCTATC 540

Qy       1222 TGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGG 1281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGTGGACCCCAACGTACCAGTGGGCTCCTGCACCAAGGACTATGTGGAGATCAACGGGG 600

Qy       1282 AGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAG 1341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTGAGCAGCAACAGCAGCAAGATTACAG 660

Qy       1342 TCCACTTCCATTCTGATCACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCT 1401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCCACTTCCATTCTGATCACTCGTACACGGACACCGGGTTCCTAGCTGAGTACCTCTCCT 720

Qy       1402 ACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAGA 1443
              |||||||||| ||||||||||||||||||||||||||||||
Db        721 ACGACTCCAAAGACCCGTGCCCAGGGATGTTCATGTGCAAGA 762


RESULT 9
CB208872
LOCUS       CB208872                874 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AGENCOURT_11350695 NIH_MGC_164 Mus musculus cDNA clone
            IMAGE:30241011 5', mRNA sequence.
ACCESSION   CB208872
VERSION     CB208872.2  GI:29133712
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     On Feb 4, 2003 this sequence version replaced gi:28250435.
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM0308  row: g  column: 04
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30241011"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_164"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Non-normalized full-length enriched library from
                     pooled mouse embryonic limb, maxilla and mandible, day
                     10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
                     Cloned directionally, priming method: Oligo-dT.  cDNA
                     enrichment: >1k bp, Average insert size 1.8k bp. Priming
                     sequence: 5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T) 3'. Tissue
                     contributed by, David Rowe. Library constructed by ResGen,
                     Invitrogen Corp."
BASE COUNT      193 a    230 c    260 g    190 t      1 others
ORIGIN

  Query Match             24.4%;  Score 759.4;  DB 14;  Length 874;
  Best Local Similarity   98.9%;  Pred. No. 9.8e-139;
  Matches  785;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    2;

Qy       1387 CTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAGACTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATGTGCAAGACTG 60

Qy       1447 GACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTATAGTG 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCGGATTATAGTG 120

Qy       1507 ATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGTTCTGCA 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAACCAGTTCTGCA 180

Qy       1567 AGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACGAGGAGG 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGTGACGAGGAGG 240

Qy       1627 GCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCC 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCC 300

Qy       1687 AGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCG 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCG 360
```

```
Qy       1747 TGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTGAGCA 1806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTGAGCA 420

Qy       1807 AGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACT 1866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACT 480

Qy       1867 GTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGG 1926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGG 540

Qy       1927 ACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTGTGTG 1986
             |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        541 ACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGNGCCAGGGCCACTTGTGTG 600

Qy       1987 GGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACA 2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACA 660

Qy       2047 AAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGA 2106
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        661 AAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACC-GA 719

Qy       2107 GCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCAT-CACCCACCCT 2165
             |||||||||||||||||||||||||||||||| |||| ||||||||| | ||||||| |
Db        720 GCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAACTCATACGTATCATCCGCCCACCGT 779

Qy       2166 TCCTTCAATGATTT 2179
             ||||||||| ||
Db        780 TCCTTCAATGGATT 793
```

RESULT 10
BI664081

```
LOCUS       BI664081                798 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603289385F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5323485 5',
            mRNA sequence.
ACCESSION   BI664081
VERSION     BI664081.1  GI:15578314
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11820  row: i  column: 22
            High quality sequence stop: 771.
FEATURES             Location/Qualifiers
     source          1. .798
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5323485"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam6"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      167 a    223 c    240 g    168 t
ORIGIN
```

```
  Query Match             24.1%;  Score 747.2;  DB 12;  Length 798;
  Best Local Similarity   98.6%;  Pred. No. 2.5e-136;
  Matches  785;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    3;

Qy         16 CCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAACCATGGGTAGCAATC 75
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCGGAGGGACCACGCGTCTGAGACCGGCGATCGGACCGCCAAAACCATGGGTAGCAATC 60

Qy         76 GGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGACTCAAGTACAACTCCC 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGGCCGCAAGGCCGGAGGGGGCTCTCAGGACTTCGGCGCGGGACTCAAGTACAACTCCC 120

Qy        136 GGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGCCTGCGAACAATGCCA 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCTAGAGAACATGAATGGCTTTGAGGAGGGTGTGGAGTTCCTGCCTGCGAACAATGCCA 180

Qy        196 AGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGGCAGTGCTGTTCAGCT 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGAAAGTGGAGAAGCGAGGCCCCAGGCGCTGGGTGGTGCTGGTGGCAGTGCTGTTCAGCT 240

Qy        256 TCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCCATTATCGGAATGTGC 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCCTCTTGCTCTCCCTCATGGCTGGCTTGCTGGTGTGGCACTTCCATTATCGGAATGTGC 300

Qy        316 GGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGATCTTTCTGGATGCGT 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGGTTCAAAAAGTCTTCAATGGCCATCTGAGGATCACAAATGAGATCTTTCTGGATGCGT 360

Qy        376 ATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGGTGAAGGAGGCGCTGA 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATGAGAACTCCACCTCCACAGAGTTTATCAGCCTGGCCAGCCAGGTGAAGGAGGCGCTGA 420

Qy        436 AGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGAAGTCGGCTGTAACTG 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGCTGTACAATGAAGTCCCTGTCCTGGGTCCCTACCACAAGAAGTCGGCTGTAACTG 480

Qy        496 CCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCAGCATCCCCCCACACC 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCTTCAGTGAGGGCAGTGTCATCGCCTACTACTGGTCAGAGTTCAGCATCCCCCCACACC 540

Qy        556 TGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAACATTGCCACCCCGAG 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGCAGAAGAGGTTGATCGCGCCATGGCTGTGGAGCGAGTTGTAACATTGCCACCCCGAG 600

Qy        616 CACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCCCCATTGACCCCAGAA 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        601 CACGGGCACTGAAATCCTTCGTGCTAACATCTGTGGTGGCCTTCCCCATTGAGCCCAGAA 660

Qy        676 TGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATG-CCCATGGTGCAGCA 734
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        661 TGCTGCAGAGGACTCAGGACAACAGCTGCAGTTTTGCCCTGCATGCCCCATGGTGCAGCA 720

Qy        735 GTG-ACACGCTTCACTACCCCTGGCTTCCCCAACAGTCCCTACCCGGCGCATGCCCGCTG 793
              ||| |||||||||| ||||||||| || |||||||||||||||||| ||||  ||| ||
Db        721 GTGAACACGCTTCAGTACCCCTGGGTT-CCCAACAGTCCCTACCCGGGGCATTGCCGTTG 779

Qy        794 CCAGTGGGTCCTGCGG 809
              ||||||||||||||||
Db        780 GCAGTGGGTCCTGCGG 795
```

RESULT 11
BY756896

```
LOCUS       BY756896                746 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY756896 RIKEN full-length enriched, blastocyst Mus musculus cDNA
            clone I1C0035F06 3', mRNA sequence.
ACCESSION   BY756896
```

VERSION BY756896.1 GI:27190109
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 746)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani;K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .746
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I1C0035F06"
/dev_stage="blastocyst"
/clone_lib="RIKEN full-length enriched, blastocyst"
BASE COUNT 172 a 193 c 224 g 155 t 2 others
ORIGIN

Query Match 23.8%; Score 739.2; DB 14; Length 746;
Best Local Similarity 99.3%; Pred. No. 9.3e-135;
Matches 741; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       2343 CTGATCCTGCAGAAGGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGGACCTCATG 2402
             |||||| ||||| | ||||||||||||||||||||||||||||||||||  |||||||
Db          1 CTGATCNTGCAGNAAGGTGAGATCCGTGTCATCAACCAGACCACCTGTGAGACCCTCATG 60

Qy       2403 CCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCC 2462
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCC 120

Qy       2463 TGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAG 2522
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGCCAGGGTGACTCTGGTGGCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAG 180

Qy       2523 GCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACACA 2582
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTGGTGTGGTGAGCTGGGGTGAAGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACACA 240

Qy       2583 AGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGACAG 2642
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGGCTCCCTGTAGTTCGGGACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGACAG 300

Qy       2643 ACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAGGA 2702
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACAGCCGACCACAAACACCCACAGGGATGCCCGACATGCACACCTGGATACAGGAGAGGA 360

Qy       2703 ACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACTGC 2762
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACACTGACGACATTTATGCTGTGGCCTCCCCCCCCCCAACACAACCCAGACTGTGAACTGC 420

Qy       2763 ATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAACTT 2822
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATCCTTAGGACTCAGAGTTCTTCCAAAGTGGGACCCCCTCAAGAGTTGGAGAGAGAACTT 480

Qy       2823 GCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCCCT 2882
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCGTGCTAGCGGCCCAGCCTGGGGGCAAGGGTTTGATGGCAGCCTTCCCCCTCTAGCCCT 540

Qy       2883 GAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCCGG 2942
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAGCTGGGTGAAGATGATGCTGTCCCGGAGAGCTGCTTCCAACTGTCATTGAGCTCCCGG 600

Qy       2943 GAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGAAC 3002
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAGCCCTATGGGAGGAGGGGCTCAGGGTCACTCTTTTCAGGAAGCGCCAGCCCTAGGAAC 660

Qy       3003 CCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTATTT 3062
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCCAGAAAAGAGTGGTACCTAAGGCTGAAATTGTTTTGCTGTTGCCAGGGGTGGGTATTT 720

Qy       3063 GAGAGTAAAACATTTTATTTCTTTTT 3088
             |||||||||||||||||||||||||||
Db        721 GAGAGTAAAACATTTTATTTCTTTTT 746
```

RESULT 12
BI416218
LOCUS BI416218 846 bp mRNA linear EST 14-AUG-2001
DEFINITION 602989878F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5145762 5',

```
            mRNA sequence.
ACCESSION   BI416218
VERSION     BI416218.1  GI:15177141
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11359  row: l  column: 19
            High quality sequence start: 20
            High quality sequence stop: 793.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5145762"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      178 a    224 c    262 g    182 t
ORIGIN

  Query Match             23.8%;  Score 739.2;  DB 12;  Length 846;
  Best Local Similarity   97.1%;  Pred. No. 9e-135;
  Matches  816;  Conservative    0;  Mismatches   18;  Indels    6;  Gaps    6;

Qy       1687 AGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCG 1746
             |||   | |   | |  |   |||||||||||||||||||||||||||||||||||||||
Db          1 AGATCCGCACGAGTAGAGGACACTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCG 60

Qy       1747 TGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTCTGTCTGAGCA 1806
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         61 TGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGGCTCTGTCTGAGCA 120

Qy       1807 AGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACT 1866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACT 180

Qy       1867 GTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGG 1926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGG 240

Qy       1927 ACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTGTGTG 1986
             |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        241 ACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGCCA-GGCCACTTGTGTG 299

Qy       1987 GGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACA 2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 GGGCCTCGCTCATCTCTCCTGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACA 359

Qy       2047 AAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGA 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 AAAATTTCAAGTACTCAGACTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGA 419

Qy       2107 GCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCCACCCTT 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 GCAAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAAACGTATCATCACCCACCCTT 479

Qy       2167 CCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGTCGGTGG 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 CCTTCAATGATTTCACCTTCGACTATGACATCGCCTTGCTGGAGCTGGAGAAGTCGGTGG 539

Qy       2227 AGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCCCTGCTG 2286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AGTACAGCACCGTCGTGCGCCCCATCTGCCTGCCTGATGCTACCCATGTCTTCCCTGCTG 599

Qy       2287 GC-AAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGT-ACCGGAGCGCT 2344
             || ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        600 GCAAAGGCCATCTGGGTCACAGGCTGGGGGCACACAAAAGAGGGAGGTCACCGGAGCGCT 659

Qy       2345 GATCCTGCAGAAGGGTGAGATCCGTGTCATC-AACCAGACCACCTGTGAGGACCTCATGC 2403
             |||||||||||||||||||||||||||||||| ||||||||| ||||||||||||||||
Db        660 GATCCTGCAGAAGGGTGAGATCCGTGTCATCAAACCAGACCGACTGTGAGGACCTCATGC 719

Qy       2404 CGCAGCAGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCCT 2463
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 CGCAG-AGATCACCCCACGAATGATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCCT 778

Qy       2464 GCCAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGG 2523
             | |||||||||||||||||||||||||||||||| ||||| ||||||||||||||||
Db        779 G-CAGGGTGACTCTGGTGGCCCCCTTGTCAAGCGCGGAGCAAGATTGGCGAATGTTCCAGG 837


RESULT 13
BF301347
LOCUS       BF301347                1116 bp    mRNA    linear   EST 21-NOV-2000
DEFINITION  602029796F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4165103 5',
            mRNA sequence.
ACCESSION   BF301347
VERSION     BF301347.1  GI:11247870
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1116)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9451  row: o  column: 24
            High quality sequence stop: 686.
FEATURES             Location/Qualifiers
     source          1. .1116
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4165103"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_SG2"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
```

NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 325 a 315 c 302 g 174 t
ORIGIN

Query Match 23.7%; Score 735.8; DB 10; Length 1116;
Best Local Similarity 92.1%; Pred. No. 3.9e-134;
Matches 800; Conservative 0; Mismatches 62; Indels 7; Gaps 2;

```
Qy       1137 CACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTG 1196
Db             1 CACTACCCGCCCAACATCAACTGCACATGGAATATCAAGGTGCCCAACAACCGGAACGTG 60

Qy       1197 AAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACC 1256
Db            61 AAGGTGCGCTTCAAACTCTTCTATCTGGTGGACCCCAACGTACCAGTGGGCTCCTGCACC 120

Qy       1257 AAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTG 1316
Db            121 AAGGACTATGTGGAGATCAACGGGGAGAAGTACTGCGGTGAGAGGTCCCAGTTTGTGGTG 180

Qy       1317 AGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACC 1376
Db            181 AGCAGCAACAGCAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACACGGACACC 240

Qy       1377 GGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATG 1436
Db            241 GGGTTCCTAGCTGAGTACCTCTCCTACGACTCCAACGACCCGTGCCCAGGGATGTTCATG 300

Qy       1437 TGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGCCCG 1496
Db            301 TGCAAGACTGGACGGTGCATCCGAAAGGAACTGCGCTGCGACGGCTGGGCAGACTGACCG 360

Qy       1497 GATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAAC 1556
Db            361 GATTATAGTGATGAGCGTTACTGCCGATGCAATGCCACCCACCAGTTCACGTGCAAAAAC 420

Qy       1557 CAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGT 1616
Db            421 CAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGACAGTGTCAACGACTGTGGGGACGGAAGT 480

Qy       1617 GACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTC 1676
Db            481 GACGAGGAGGGCTGCAGCTGTCCTGCTGGGAGTTTCAAGTGTTCCAATGGGAAGTGTCTC 540

Qy       1677 CCTCAGAGCCAGAAGTGTAATGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCA 1736
Db            541 CCTCAGAGCCAGAAGTGTACTGGGAAGGACAACTGTGGAGATGGGTCTGACGAGGCTTCA 600

Qy       1737 TGTGACAGCGTGAATGTCGTCTCTTGCACCAAATATACCTACCGCTGCCAAAATGGCCTC 1796
Db            601 TGTGACAGCGTGAATGTCGTCTCTTGCACCAGATATACCTACCGATGCCAAAAAGGCCTC 660

Qy       1797 TGTCTGAGCAAGGGCAACCCTGAGTGTGATGGGAAGACGGACTGTAGCGATGGCTCCGAT 1856
Db            661 TGTCTGAGCAAGGGCCACCCTGAGTGTGAGGGGAAGACCGACTGTACGAATGGCTCCGCA 720

Qy       1857 G---AGAAAAACTGTGACTGTGGGCTGCGATCCTTTACCAAACAGGCTCGCGTGGTTGGT 1913
Db            721 GGACAAACACACTGGGACTGTGGGCAGCGATCCTTTACCAAACAGGATCAGGTGACAGGT 780

Qy       1914 GGCACGAATGCG----GACGAGGGCGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGG 1969
Db            781 GGGAACGAATTGCAGACAAAGGGACAAAGTGCCCGTGCACGAGAGCCTCCACGGCCTGTA 840

Qy       1970 CCAGGGCCACTTGTGTGGGGCCTCGCTCA 1998
Db            841 GAACAGGCCAATGAGTGGAACCCCTAGCA 869
```

RESULT 14
BI646380

```
LOCUS       BI646380                786 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603276244F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5316539 5',
            mRNA sequence.
ACCESSION   BI646380
VERSION     BI646380.1  GI:15560616
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11802  row: h  column: 12
            High quality sequence stop: 764.
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5316539"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam3"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
BASE COUNT      177 a    203 c    235 g    171 t
ORIGIN
```

Query Match 23.6%; Score 732; DB 12; Length 786;
Best Local Similarity 98.7%; Pred. No. 2.4e-133;
Matches 780; Conservative 0; Mismatches 5; Indels 5; Gaps 4;

```
Qy       1409 CAACGACCCGTGCCCAGGGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACT 1468
Db             1 CAACGACCCGTGCCCA-GGATGTTCATGTGCAAGACTGGACGGTGCATCCGAAAGGAACT 59

Qy       1469 GCGCTGCGACGGCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAA 1528
Db            60 GCGCTGCGACGGCTGGGCAGACTGCCCGGATTATAGTGATGAGCGTTACTGCCGATGCAA 119

Qy       1529 TGCCACCCACCAGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGA 1588
Db           120 TGCCACCCACCAGTTCACGTGCAAAAACCAGTTCTGCAAGCCCCTCTTCTGGGTCTGTGA 179

Qy       1589 CAGTGTCAACGACTGTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAG 1648
Db           180 CAGTGTCAACGACTGTGGGGACGGAAGTGACGAGGAGGGCTGCAGCTGTCCTGCTGGGAG 239

Qy       1649 TTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAA 1708
Db           240 TTTCAAGTGTTCCAATGGGAAGTGTCTCCCTCAGAGCCAGAAGTGTAATGGGAAGGACAA 299

Qy       1709 CTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAA 1768
Db           300 CTGTGGAGATGGGTCTGACGAGGCTTCATGTGACAGCGTGAATGTCGTCTCTTGCACCAA 359
```

```
Qy       1769 ATATACCTACCGCTGCCAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGG 1828
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 ATATACCTACCGCTGCCAAAATGGCCTCTGTCTGAGCAAGGGCAACCCTGAGTGTGATGG 419

Qy       1829 GAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTT 1888
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 GAAGACGGACTGTAGCGATGGCTCCGATGAGAAAAACTGTGACTGTGGGCTGCGATCCTT 479

Qy       1889 TACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCA 1948
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 TACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTGGCA 539

Qy       1949 GGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGA 2008
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 GGTGAGCCTCCACGCCCTGGGGCCAGGCCACTTGTGTGGGGCCTCGCTCATCTCTCCTGA 599

Qy       2009 CTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTA 2068
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 CTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGACTA 659

Qy       2069 CACG-ATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGG 2127
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 CACGAATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGGGG 719

Qy       2128 TGCAGGAGCTGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTTCG 2187
             ||||||||||||||||||||||||||||||  |||  | |||||||||  ||||||||||
Db        720 TGCAGGAGCTGAAGCTCAAACGTATCATCACCAACC--TTCTTCAATGA-TTCACCTTCG 776

Qy       2188 ACTATGACAT 2197
             ||||||||||
Db        777 ACTATGACAT 786


RESULT 15
BG915045
LOCUS       BG915045                 928 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  602814325F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4936590 5',
            mRNA sequence.
ACCESSION   BG915045
VERSION     BG915045.1  GI:14295521
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10869  row: a  column: 07
            High quality sequence stop: 793.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4936590"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam4"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and  Differentiation 7, 3-11 (1996)."
BASE COUNT      216 a    258 c    273 g    181 t
ORIGIN

  Query Match           23.2%;  Score 721;  DB 12;  Length 928;
  Best Local Similarity   93.2%;  Pred. No. 3.3e-131;
  Matches  832;  Conservative    0;  Mismatches   50;  Indels   11;  Gaps    7;

Qy       1886 CTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTG 1945
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTTTACCAAACAGGCTCGCGTGGTTGGTGGCACGAATGCGGACGAGGGCGAGTGGCCCTG 60

Qy       1946 GCAGGTGAGCCTCCACGCCCTGGGCCAGGGCCACTTGTGTGGGGCCTCGCTCATCTCTCC 2005
             |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db         61 GCAGGTGAGCCTCCACGCCCTGGGCCA-GGCCACTTGTGTGGGGCCTCGCTCATCTCTCC 119

Qy       2006 TGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGA 2065
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 TGACTGGCTGGTCTCTGCAGCTCATTGCTTTCAGGATGACAAAAATTTCAAGTACTCAGA 179

Qy       2066 CTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGG 2125
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 CTACACGATGTGGACGGCCTTCCTGGGTCTGCTGGACCAGAGCAAGCGCAGTGCCTCTGG 239

Qy       2126 GGTGCAGGAGCTGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTT 2185
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GGTGCAGGAGCTGAAGCTCAAACGTATCATCACCCACCCTTCCTTCAATGATTTCACCTT 299

Qy       2186 CGACTATGACATCGCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCG 2245
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 CGACTATGACATCGCCTTGCTGGAGCTGGAGAAGTCGGTGGAGTACAGCACCGTCGTGCG 359

Qy       2246 CCCCATCTGCCTGCCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCAC 2305
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 CCCCATCTGCCTGCCTGATGCTACCCATGTCTTCCCTGCTGGCAAGGCCATCTGGGTCAC 419

Qy       2306 AGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGAT 2365
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 AGGCTGGGGGCACACAAAAGAGGGAGGTACCGGAGCGCTGATCCTGCAGAAGGGTGAGAT 479

Qy       2366 CCGTGTCATCAACCAGACCACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAAT 2425
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 CCGTGTCATCAACCAGACCACCTGTGAGGACCTCATGCCGCAGCAGATCACCCCACGAAT 539

Qy       2426 GATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGT--GGC 2483
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||  | |
Db        540 GATGTGTGTGGGTTTCCTCAGTGGGGGTGTGGACTCCTGCCAGGGTGACTCTGGTTGGCC 599

Qy       2484 CCCTTGTCAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGT 2543
             || |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 CCTTGTCAAAGCGCGGAGAAAGATGGGCGAATGTTCCAGGCTGGTGTGGTGAGCTGGGGT 659

Qy       2544 GAAGGCTGCGCTCAGAGGAACAAG-CCAGGCGTGTACA-CAAGGCTCCCTGTAGTTCGGG 2601
             |||||||||||||||||||||||||| |||||||||||| |||| ||||||||||||||||
Db        660 GAAGGCTGCGCTCAGAGGAACAAGCCCAGGCGTGTACACCAAGGGTCCCTGTAGTTCGGG 719

Qy       2602 ACTGGATCAAAGAGCACACTGGGGTATAGCAGCATGGA-CAGACAGCCGACCACAAACAC 2660
             | |   |||| ||||||||||||||||||||||||| |||||||       ||||||
Db        720 AACTGGATCAAGAACACACTGGGGTATAGCAGCATGGACCAGACAGCGACAACAAAACAC 779

Qy       2661 CCACAGGGATGCCCGACATGCACACCTGGAT---ACAGGAGAGGAACACTGACGACATTT 2717
             |||||||||||||| ||||||||||||||   |   |   | | ||||||||||||
Db        780 CCACAGGGATGCCCGAAATGCACACCTGGATTACAGGAGCAGGAACCACTGACGACATCA 839

Qy       2718 ATGCTGTGGCCTCCCCCCCCCAACACAACCCAGACTGTGAACTGCATCCTTAG 2770
             |||||| ||| |||||  | |  | |  ||||||||||   ||| ||
Db        840 TGCTTGTGGCATCCACCCCC--AAAACAACACAACTGTGAACTGGTACCTAAG 890
```

Search completed: February 21, 2004, 07:12:09
Job time : 6230 secs